קידוד אותות דיבור בקצבים נמוכים מאד באמצעות פירוק למאורעות זמניים

(Temporal Decomposition)

חיבור על מחקר

מוגש לשם מילוי חלקי של הדרישות לקבלת התואר
מגיסטר למדעים בהנדסת חשמל

סלבה שכטמן

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
אב תשס״ד חיפה יולי 2004

לנדושה וטליק האהובים

המחקר נעשה בהנחיית פרופ׳ דוד מלאך בפקולטה להנדסת חשמל.

תודתי העמוקה לפרופ׳ דוד מלאך על הנחייתו המסורה והיסודית, מעורבותו הרבה והדיונים המועילים.

תודה לכל צוות המעבדה לעיבוד אותות ותמונות: נמרוד פלג, יאיר משה, זיוה אבני, אבי רוזן ותמרה גבירץ על העזרה והתמיכה הטכנית.

אני מודה ל גיא נרקיס, אסף הרגיל וצחי זהבי על מימוש ואספקת תוצאות הרצה של המערכת ששימשה כנקודת ההשוואה העיקרית לאלגוריתם שפותח בעבודה זו.

ברצוני להודות לאשתי נדיה על אהבתה, תמיכתה וסבלנותה האינסופית, וכן להוריי ולהוריה אשר תרמו רבות להצלחתי.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

תוכן העניינים

תוכן העניינים i

רשימת טבלאות ואיורים v

תקציר 1

רשימת סמלים וקיצורים 3

פרק 1. מבוא 6

פרק 2. רקע למערכת קידוד הדיבור 10

2.1 מבנה אות הדיבור 10
2.2 מבנה מקודד דיבור פרמטרי 11
2.3 מודל ה-LPC של אות הדיבור ופרמטרי ייצוג של המעטפת הספקטרלית 11
2.3.1 המודל 11
2.3.2 שערוך המעטפת הספקטרלית באמצעות חיזוי לינארי 13
2.3.3 ייצוג מסנן ה-LPC בעזרת LSF 14
2.4 פונקצית עוות של מקדמי מסנן ה-LPC 15
2.4.1 מדד ה-Log-Spectral Distortion (LSD) 16
2.4.2 מדדים המבוססים על שגיאה ריבועית משוקללת 16
2.5 עירור מעורב המשולב עם מערכת חיזוי לינארי (MELP) 18

פרק 3. שיטות לקידוד פרמטרים ספקטרליים 21

3.1 מבוא 21
3.2 קידוד פרמטרים ספקטרליים 21
3.3 ניצול תלות זמנית עם השהיה קצרה 22
3.4 שיטות המתבססות על דילוג מסגרות 23
3.5 כימות מטריצי (MQ) 23
3.6 כימות סגמנטים (SegQ) 25
3.7 ייצוג סגמנטים עם מסגרת אנליזה בסיסית באורך משתנה 27
3.8 סיכום 29

פרק 4. מודל הפירוק למאורעות זמניים (Temporal Decomposition) 30

4.1 מבוא 30
4.2 המודל הכללי של TD 30
4.2.1 תיאור כללי 30
4.2.2 מציאת מטריצת וקטורי המטרה 33
4.3 שיטות פירוק המבוססות על SVD 33
4.4 שיטות מודרניות לביצוע TD בהנחת סגמנטציה התחלתית 36
4.4.1 הגבלת התמך של פונקציות במאורעות ע"י כופלי לגראנז' 37
4.4.2 אילוץ קשיח של תמך המאורעות (RTD) 39
4.5 פירוק TD אופטימלי מבוסס RTD (ORTD) 43
4.5.1 תיאור כללי 43
4.5.2 חפיפה בין הבלוקים ותנאי קצה 45
4.6 סיכום 46

פרק 5. קידוד מעטפת דיבור באמצעות DW-SORTeD 47

5.1 מבוא 47
5.2 בחירת הקריטריונים לקירוב המעטפת הספקטרלית 48
5.2.1 כללי 48
5.2.2 מדד ה-G-WSE המסונן 48
5.3 פירוק RTD אופטימלי (ORTD) עם קריטריון של WSE דינמי (DW-ORTD) 50
5.3.1 מודיפיקציות לשלב של מציאת פונקציות המאורעות 50
5.3.2 מודיפיקציות לשלב של עידון וקטורי המטרה 50
5.4 אלגוריתם תת-אופטימלי למציאת פונקציות המאורע. 53
5.4.1 מבוא 53
5.4.2 תיאור כללי 54
5.4.3 עדכון של הבלוק לאנליזה 55
5.4.4 קביעה של סגמנטציה התחלתית 56
5.4.5 אלגוריתם תת-אופטימלי לשיפור הסגמנטציה ההתחלתית 59
5.4.6 השוואת הסיבוכיות של האלגוריתמים ORTD ו-SORTeD 62
5.4.7 דרכים לחישוב פונקציות מאורע רגעיות 64
5.5 קוונטיזציה של פרמטרי ה-TD 70
5.6 סיכום 72

פרק 6. בחינת ביצועים של אלגוריתם ה-DW-SORTeD 73

6.1 מבוא ... 73
6.2 הערכת הביצועים של מודל ה-TD ... 74
6.2.1 הקדמה ... 74
6.2.2 ORTD לעומת DW-ORTD עם משקלות שונים ... 74
6.2.3 בחינת פרמטרים ל-DW-SORTeD ... 76
6.2.4 DW-ORTD לעומת DW-SORTeD ... 77
6.2.5 הוספת האילוצים על פונקציות המאורעות ב-DW-SORTeD ... 78
6.2.6 בחינת הביצועים כתלות במספר האיטרציות ב-DW-SORTeD ... 79
6.3 הערכת הביצועים של המודל עם קוונטיזציה ... 80
6.3.1 בחינת ספר קוד לוקטורי המטרה ... 81
6.3.2 בחינת הקידוד של פונקציות המאורע ... 81
6.3.3 בחינת הביצועים של DW-SORTeD בקצבים שונים ובהשהיות שונות ... 82
6.4 סיכום ... 85

פרק 7. דחיסת פרמטרי העירור באמצעות אלגוריתם DW-SORTeD ... 86

7.1 מבוא ... 86
7.2 התאמת ה-RTD המאולץ לביצוע ה-TD לוקטורי פרמטרים מנורמלים ... 87
7.2.1 מבוא ... 87
7.2.2 RTD עם אילוץ השלמה לאחד תחת התמרה אפינית (affine transform) ... 88
7.3 מודל לביצוע ה-TD למספר פרמטרי העירור ... 89
7.3.1 תיאור כללי ... 89
7.3.2 עיבוד מקדים של פרמטרי העירור ... 90
7.3.3 נרמול ... 91
7.3.4 קביעת המשקל הדינמי לביצוע ה-DW-ORTD ... 93
7.3.5 התאמת ה-DW-SORTeD לפרמטרי העירור ... 95
7.4 בחינת ביצועים של קידוד פרמטרי העירור באמצעות ה-TD ... 95
7.4.1 כללי ... 95
7.4.2 דוגמת הרצה ... 95

פרק 8. מקודד דיבור בקצבים נמוכים מאד המבוסס MELP ו-DW-SORTeD ... 99

8.1 הקדמה ... 99
8.2 מקודד מבוסס MELP ל-600 bps ... 99
8.2.1 כללי ... 99

8.2.2 קוונטיזציה של תבנית ה-voicing (קוליות). 100
8.2.3 הקצאת הסיביות למקודד 600-700 bps 102
8.2.4 הקצאת הסיביות למקודד 800-880 bps 104
8.3 בחינת ביצועים 105
8.3.1 סכימאות קידוד להשוואה 105
8.3.2 תוצאות ההשוואה 106
8.3.3 דיון ומסקנות 107
פרק 9. סיכום והצעות להמשך המחקר 109
9.1 סיכום 109
9.2 הצעות להמשך מחקר 111
נספח א' 112
נספח ב' 114
ביבליוגרפיה 116

רשימת טבלאות ואיורים

רשימת טבלאות

טבלה 2-א. איכות שקופה של מעטפת ספקטרלית של הדיבור, כפי שמוגדר ע״י מדד ה-LSD. 16

טבלה 2-ב. הקצאת סיביות למקודד דיבור MELP עבור מסגרת בודדת 20

טבלה 5-א. אפשרויות שנבחנו לקביעת הסגמנטציה ההתחלתית עבור אלגוריתם ה-SORTeD 58

טבלה 5-ב. האלגוריתם התת-אופטימלי למציאת הסגמנטציה לבלוק – הפעלה ראשונה 60

טבלה 5-ג. האלגוריתם התת-אופטימלי למציאת הסגמנטציה של בלוק – הפעלה כלשהי 61

טבלה 5-ד. סיבוכיות של האלגוריתם האופטימלי והתת-אופטימלי 63

טבלה 6-א. עיוותים ספקטרליים (ה-LSD הממוצע ואחוזי החריגות), אשר מתקבלים עבור שקלולים שונים של מדד השגיאה באלגוריתם ה-DW-ORTD.. 76

טבלה 6-ב. תוצאות הניסויים לבחינת התצורות של תנאי ההתחלה של ה-DW-SORTeD. 76

טבלה 6-ג. עיוותים ספקטרליים (ה-LSD הממוצע ואחוזי החריגות), אשר מתקבלים עבור אלגוריתמי ה-DW-ORTD ו-DW-SORTeD.. 77

טבלה 6-ד. ביצועים של האלגוריתם ה-DW-SORTeD המאולץ בהשוואה לאלגוריתם ללא אילוצים על פונקציות המאורע. 79

טבלה 6-ה. ביצועים של תצורות שונות של האלגוריתם ה-DW-SORTeD המאולץ...................... 79

טבלה 6-ו. ביצועים של ספרי קוד שונים בקוונטיזציה של וקטורי ה-LSF בסכימה של Slpit-VQ. .. 81

טבלה 6-ז. הביצועים של ה-DW-SORTeD ללא קוונטיזציה של וקטורי המטרה......................... 82

טבלה 6-ח. הערכת הביצועים של מקודד ה-DW-SORTeD עבור המעטפת הספקטרלית של הדיבור. 84

טבלה 7-א. ביצועים של קוונוט פרמטרי העירור באמצעות ה-TD... 97

טבלה 8-א. טבלת קוונטיזציה של תבנית ה-voicing. 102

טבלה 8-ב. הקצאת סיביות למקודד מבוסס MELP. ההקצאה היא עבור קידוד של 11 מסגרות. .. 103

טבלה 8-ג. הקצאת סיביות למקודד מבוסס MELP. ההקצאה היא עבור קידוד של 7 מסגרות...... 104

טבלה 8-ד. הערכת ביצועים של מקודדי דיבור בקצבים נמוכים מאד. 106

רשימת איורים

איור 2-א מערכת קידוד דיבור פרמטרי כללית 11

איור 2-ב. מערכת קידוד דיבור LPC הבסיסית........ 12

איור 3-א. PVQ קדמי-אחורי לקידוד וקטורי ה-LSF בתקן MELP-1200........ 22

איור 3-ב מציאת J סגמנטים ע״י תכנות דינמי 26

איור 3-ג. . דיאגרמת הסריג עבור האלגוריתם הדינמי.. 29

איור 4-א. סימונים עבור מודל הפירוק למאורעות זמניים (Temporal Decomposition)........ 31

איור 4-ב. שלבי ה-TD הכללי........ 32

איור 4-ג. פירוק ה-TD הכללי. 32

איור 4-ד. פונקצית משקל עבור המאורע. 37

איור 4-ה. . הצורה האופיינית של פונקצית מאורע התחלתית........ 38

איור 4-ו. פירוק ה-TD המצומצם (RTD). 39

איור 4-ז. פירוק RTD עם פונקציות מאורע אופטימליות (לא מאולצות). 40

איור 4-ח. . פיזור של פונקציות מאורע רגעיות אופטימליות עבור מודל ה-RTD. 41

איור 4-ט. . פירוק RTD עם פונקציות מאורע המשלימות ל-1........ 42

איור 4-י. דוגמה לדיאגרמת trellis המשמשת לחיפוש אתר הסגמנטציה האופטימלית ב-ORTD.... 44

איור 4-יא. החפיפה בין החוצצים ב-ORTD........ 45

איור 5-א. יכולת התאמה למדד ה-LSD של שגיאות ריבועיות משוקללות........ 49

איור 5-ב. אלגוריתם ה-RTD התת-אופטימלי (SORTeD). תיאור כללי 55

איור 5-ג. חפיפה בין הבלוקים השומרת על קצב קבוע של המאורעות לשנייה. 56

איור 5-ד. פירוק RTD עם פונקציות מאורע המשלימות לאחד, אי שליליות וממורכזות.. 66

איור 5-ה. פירוק RTD עם פונקציות מאורע המשלימות לאחד, אי שליליות ממורכזות ומונוטוניות.. 68

איור 5-ו. אילוץ מונוטוניות משופר........ 70

איור 6-א. עיוותים ספקטרליים (ה-LSD הממוצע), אשר מתקבלים עבור שקלולים שונים של מדד השגיאה באלגוריתם ה-DW-ORTD........ 75

איור 6-ב. עיוותים ספקטרליים (ה-LSD הממוצע), אשר מתקבלים עבור אלגוריתמי ה-DW-ORTD ו-DW-SORTeD. 78

איור 6-ג. שיפור של הביצועים של DW-SORTeD בתלות במספר האיטרציות שלו. 80

איור 6-ד. הערכת הביצועים של מקודד ה-DW-SORTeD עבור המעטפת הספקטרלית של הדיבור ע״י מדד ה-LSD הממוצע........ 84

איור 6-ה. הערכת הביצועים של מקודד ה-DW-SORTeD עבור המעטפת הספקטרלית של הדיבור בשילוב עם קידוד העירור התקני של תקן MELP-2400 באמצעות ציון ה-PESQ. 85
איור 7-א. מסלול ההשתנות של האנרגיה וה-pitch לאורך זמן 87
איור 7-ב פעולת ה-TD לוקטורי פרמטרי עירור 90
איור 7-ג. עקיבה אחר הערך המכסימלי והמינימלי של האנרגיה, המשמשת לשם נורמליזציה של וקטור פרמטרי העירור 92
איור 7-ד. עקיבה אחר הערך הממוצע של תדר ה-pitch, המשמשת לנורמליזציה של וקטור פרמטרי העירור.. 92
איור 7-ה. משקלים דינמיים לאנרגיה עבור קידוד משותף של תדר ה-pitch והאנרגיה באמצעות DW-SORTeD.. 94
איור 7-ו. משקלים דינמיים לתדר ה-pitch עבור קידוד משותף של תדר ה-pitch והאנרגיה באמצעות DW-SORTeD... 94
איור 7-ז. הפעלת ה-DW-SORTeD על פרמטרי ה-pitch והאנרגיה במשותף.............................. 96
איור 8-א. סכימה מלבנית של המקודד 600 bps.. 100
איור 8-ב, הערכת ביצועים של מקודדי דיבור בקצבים נמוכים מאד.. 107
איור 8-ג. פגיעה באיכות הדיבור כתוצאה מהורדת קצב הסיביות עבור אלגוריתמי קידוד שונים. .. 108

תקציר

עבודה זו מתמקדת בקידוד דיבור לקצבים נמוכים ביותר באמצעות מודל לייצוג המעטפת הספקטרלית של אות הדיבור, הקרויה Temporal Decomposition (TD). מוצע אלגוריתם לייצוג וקידוד יעיל של פרמטרי המעטפת הספקטרלית אשר מתוארים בעזרת Line Spectral Frequencies (LSF). האלגוריתם המוצע מתבסס על מודיפיקציה של מודל ה-TD, הקרוי Optimized Restricted Temporal Decomposition (ORTD), אשר הוצעה ב-[43,44]. האלגוריתם המוצע מכונה Dynamically Weighted Sub-optimal Temporal Decomposition, או בקיצור DW-SORTeD, פותח ע"י ההכללה של קריטריון השגיאה ומשיג הורדת העומס החישובי של ה-ORTD. שינוי של קריטריון השגיאה באלגוריתם זה מאפשר להגיע למינימיזציה של קריטריוני השגיאה הנמצאים בקורלציה גבוהה עם מודל השמיעה האנושית, וכך להגיע להתאמה טובה יותר למעטפת הספקטרלית במובן של ה-Log-spectral-distortion (LSD). האלגוריתם הינו בעל קצב שידור קבוע וכן עומס חישובי שמתאים למערכות זמן-אמת מעשיות.

אלגוריתם ה-DW-SORTeD מאפשר לקודד את המעטפת הספקטרלית המיוצגת ע"י וקטורי ה-LSF ב-300-380 סיביות לשנייה, כאשר ההשהיה האלגוריתמית של המקודד נקבעה להיות 160 או 250 מילישניות. ההתאמה המושגת של המעטפת הספקטרלית באלגוריתם זה היא 2.1-2.25 dB, עפ"י מדד ה-LSD. קידוד המעטפת הספקטרלית באמצעות DW-SORTeD שולב עם העירור התקני עפ"י תקן ה-MELP-2400 [6] (Mixed Excitation Linear Prediction), כדי לאפשר הערכת הדגרדציה היחסית של אות המוצא המתקבל ב-MELP ע"י קידוד המעטפת הספקטרלית באמצעות קוונטיזציה הווקטורית עם 1111 סיביות לשנייה לעומת העיוות המתקבל עם ה-DW-SORTeD בקצבים הנידונים (300-380 סל"ש). ה-DW-SORTeD בקצבים הנקובים. תוצאות הבדיקות המעשיות הראו, כי הדגרדציה בציון ה-PESQ [60] במקרה זה היא 0.2 בלבד (מ-3.0 ל-2.8). (PESQ או ITU P.862 הנו תקן להערכת ציון ה-MOS (Mean Opinion Score) הסובייקטיבי של איכות הדיבור במוצא המקודדים.)

מקודד ה-MELP מפיק אות דיבור משוחזר שאיכותו גבוהה יותר מאשר אות הדיבור המשוחזר שמתקבל ממקודד ה-LPC-10 הקלאסי [5], המבוסס על עירור הלמים או רעש לסינתזה של אות קולי או א-קולי, בהתאמה. השיפורים העיקריים של מודל ה-MELP, לעומת מקודד ה-LPC-10, כוללים עירור מעורב של פולסים ורעש לבן, אפשרות של שימוש בפולסים לא מחזוריים עבור דיבור קולי, מסנן מסתגל לשיפור טבעיות האות הנשמע ושימוש במסנן לעיצוב הפולס. מקודד זה נבחר ב-1996 להיות תקן ה-DoD בקצב של 2400bps.

בעבודה זו הורדת קצב השידור של מקודד ה-MELP מ-2400bps ל-600bps נעשית בעיקר של ידי שימוש באלגוריתם ה-DW-SORTeD לדחיסת המעטפת הספקטרלית ופרמטרי העירור (האנרגיה וה-pitch). איכות האותות המתקבלים הניבה ציונים של 2.6-2.65 ב- PESQ, עבור המקודדים בקצבים 600-650 סיבות לשנייה, בהתאמה, עבור התצורה עם השהיה אלגוריתמית של 11 מסגרות (250 מילישניות). המערכת הדומה עם ההשהיה האלגוריתמית של 7 מסגרות (160 מילישניות) הפיקה דיבור באיכות של 2.5-2.6 (PESQ) בקצבים 610-650 סיביות לשנייה, בהתאמה. המקודד המוצע הושווה למקודד אחר בקצב דומה [57], גם הוא מבוסס על תקן ה-MELP. המקודד המוצע המבוסס על DW-SORTeD, הראה שיפור משמעותי של 0.2-0.3 (PESQ) עבור נקודות העבודה בקצבים 600-650 סיביות לשנייה, בהתאמה.

רשימת סמלים וקיצורים

רשימת סמלים

$A_p(z)$ -	פולינום המכנה של מסנן הסינטזה מסדר p
$P_{p+1}(z)$ -	פולינום למציאת תדרי LSF אי-זוגיים
$Q_{p+1}(z)$ -	פולינום למציאת תדרי LSF זוגיים
$d_{LSD}(A,\hat{A})$ -	מדד LSD בין המעטפת הספקטרלית לפני ואחרי קוונטיזציה
$\mathbf{W}(n)$ -	מטריצת משקלים בעיוות ריבועי משוקלל המתאימה למסגרת ה-n-ית.
$d_{WMSE}(x,y)$ -	עיוות ריבועי עם משקל (WMSE)
$\mathbf{Y}$ -	מטריצת הפרמטרים הספקטרליים, המכילה את וקטורי הפרמטרים ספקטרליים
$\hat{\mathbf{Y}}$ -	מטריצת הפרמטרים הספקטרליים המקורבת ע"י מודל ה-TD.
$\mathbf{y}(n)$ -	וקטור הפרמטרים הספקטרליים של המסגרת ה-n-ית
$\hat{\mathbf{y}}(n)$ -	וקטור הפרמטרים הספקטרליים המקורב ע"י מודל ה-TD.
$\mathbf{A}$ -	מטריצת וקטורי המטרה כווקטורי עמודה
$\mathbf{a}_k$ -	וקטור המטרה המתאים למאורע ה-k-י.
$\mathbf{\Phi}$ -	מטריצת פונקציות המאורע
$\varphi_k(n)$ -	ערך פונקצית המאורע ה-k-ית במסגרת ה-n-ית.
n_k -	מרכז המאורע ה-k-ית
$E(n_1,n_2)$	שגיאת ההתאמה עבור הסגמנט $[n_1,n_2)$

רשימת קיצורים

CELP	Code Excited Linear Prediction
CQI	Combined Quantization Interpolation
DW-	Dynamically Weighted
FSVQ	Finite State Vector Quantization
GCI	Glottal Closure Instant
G-WSE	Gardner Weighted Squared Error
LA	Log Area
LAR	Log Area Ratios
LPC	Linear Prediction Coding
LSD	Log Spectral Distance
LSF	Line Spectral Frequencies
MELP	Mixed Excitation Linear Prediction
MQ	Matrix Quantization
ORTD	Optimal Resctricted Temporal Decomposition
PA-WSE	Paliwal-Atal Weighted Squared Error
PESQ	Perceptual Evaluation of Speech Quality
PCM	Pulse Code Modulation
PVQ	Predicted Vector Quantization
RTD	Restricted Temporal Decomposition
SE	Squared Error
SegQ	Segment Quantization

SFTR	Spectral Feature Transition Rate
SORTeD	Sub-Optimal Restricted Temporal Decomposition
TD	Temporal Decomposition
TSQ	Trellis Segment Quantization
VLBR	Very Low Bit Rate
VQ	Vector Quantization
WMSE	Weighted Mean Squared Error
WSE	Weighted Squared Error

פרק 1

מבוא

אחת הבעיות החשובות בתחום עיבוד אותות הדיבור היא ייצוג ספרתי יעיל וחסכוני לצרכי אחסון או תמסורת. ייצוג ספרתי של אותות דיבור משמש כיום במגוון רב של שימושים כגון אחסון של אותות דיבור, מערכות מענה אוטומטיות ומערכות תקשורת צבאיות ואזרחיות הכוללות ערוצים אלחוטיים או קוויים. אחד המאפיינים החשובים של הייצוג הספרתי הוא קצב התמסורת, כלומר מספר הסיביות הנדרשות לתיאור של שנייה אחת של אות דיבור. חשיבות זו נובעת מכך שערוצי תקשורת רבים (בייחוד ערוצים אלחוטיים) הינם בעלי קיבולת מוגבלת (כלומר ניתן להעביר בהם כמות מוגבלת של מידע ספרתי לשנייה עם קצב שגיאות ממוצע רצוי). יכולת להעביר אות דיבור בקצב תמסורת נמוך מאפשרת גם ריבוב מספר רב של אותות דיבור בתוך ערוץ תמסורת אחד לצורך הורדת עלויות התמסורת (דוגמה בולטת היא טלפונים לווייניים, בהם עיקר עלות השידור הוא עלות רוחב סרט בלוויין).

אות הדיבור המקורי הוא אות אנלוגי ועבור ערוצי תקשורת רבים (כגון טלפוניה) מתכוונים, בד״כ, לאות חסום סרט ברוחב 4 kHz, הדגום בקצב של 8 kHz ברזולוציה של 8 bits/sample (תוך שימוש במקוונטים /quantizers/ לוגריתמיים). לפיכך, קצב המקור של אות הדיבור הספרתי (הידוע כאות PCM) הוא 64 kbps.

התמסורת הספרתית של אות הדיבור נעשית בעזרת מערכת קידוד המורכבת ממקודד ומפענח, כאשר המטרה של מערכת הקידוד להשתמש במינימום סיביות במוצא המקודד ולשמור על איכות טובה ככל האפשר במוצא המפענח. המקודד מקבל בכניסתו את אות הדיבור המקורי (או דגימות שלו) ומייצר סדרת סיביות בקצב נמוך (bit-stream). סדרת הסיביות (בד״כ לאחר מעבר דרך ערוץ כלשהו) משמשת אות הכניסה למפענח אשר מייצר אות דיבור משוחזר.

באופן כללי ניתן לומר, כי ישנם מספר תכונות כלליות שמאפיינות מקודד דיבור: קצב הסיביות (bit rate) , ההשהיה (delay) , הסיבוכיות (complexity) והאיכות (quality). תכונות אלו קשורות זו בזו ולעיתים שיפור של תכונה מסוימת הוא על חשבון תכונה אחרת. לכן, בהתאם לאופי השימוש יש לקבוע אילו מהתכונות הללו הן החשובות ביותר ובהתאם לכך לבחור בשיטת הקידוד הרצויה.

ישנם מספר הגדרות של איכות הדיבור, אשר משמשות לסווג מקודדי דיבור. האיכות המלאה של האות הדגום ב-4 kHz (toll quality) מוגדרת כאיכות שקופה על פני ערוץ טלפוניה תקני (רוחב סרט של 200-3200 Hz, עם SNR מעל 30 dB), אות בעל איכות תקשורת (communication quality) הינו בעל מובנות גבוהה אבל עם הפרעות הניתנות להבחנה בהשוואה לאות PCM. איכות סינטטית של

אות הדיבור מתאפיינת במובנות של מעל 80-90%. אות דיבור בעל איכות זו סובל מהפרעות ועיוותים כגון צליל "מכני" ומתכתי, זמזומים, קליקים והעדר הבחנת הדובר.

פרמטר נוסף חשוב הוא השהית המערכת. השהיות של מקודדים המשמשים במערכות טלפוניה מסחריות קוויות וסלולריות מתאפיינות בהשהיות נמוכות (50ms לכל היותר). לעומת זאת השהיות ארוכות יותר (200-300ms) ייתכנו בשימושים צבאיים בערוצי תקשורת half-duplex.

ניתן כיום למצוא מגוון רב של אלגוריתמים ותקנים (אלו של ה-DoD ו-ITU) הפועלים בקצבים 1.2-64 kbps ומתאימים לשימושים שונים. מקודדי צורת גל (Waveform Coders, המנסים לשמר את התיאור הזמני של אות הדיבור) פועלים בקצבים גבוהים יחסית (16-32 kbps). הם בעלי סיבוכיות והשהיה נמוכות יחסית ומניבים איכות מלאה (toll quality). מקודדי הדיבור לקצבים נמוכים מאד (מתחת ל-4 kbps) הם בדרך כלל פרמטריים, כלומר מבוססים על מודל שבאמצעותו ניתן לייצר אות שנשמע כמו דיבור, כאשר הפרמטרים מייצגים את מודל הדיבור. בצורה כזו ניתן להגיע ליעילות קידוד גבוהה במחיר של איכות דיבור (האיכות נעה בין איכות תקשורת לאיכות סינטטית). משפחת מקודדים היברידיים (דוגמת Code Excitation Linear CELP - Prediction) המשלבים גישה פרמטרית עם קידוד של צורת גל פועלים בטווח הקצבים 4-8 kbps. הם משיגים פשרה בין יעילות קידוד ואיכות. מקודדים אילו עובדים בהשהיות נמוכות ומניבים איכות הנעה בין איכות תקשורת עד לאיכות מלאה (toll) עבור קצבים בסביבות 8 kbps.

באחרונה הוגברה פעילות המחקר והסטנדרטיזציה של מקודדים בקצבים נמוכים ביותר (מתחת ל-2 kbps). שימושם העיקרי של מקודדים אילו (פרט לעניין אקדמי רב) הוא בתחום התקשורת הצבאית (למשל בערוצי HF בעלי קיבול נמוך וקצב שגיאות גבוה) והם מתאפיינים בהשהיה גבוהה יחסית ואיכות סינטטית (או "סינטטית משופרת").

ככלל, סיבוכיות המקודדים משתנה ביחס הפוך לקצב התמסורת. מקודדי צורת גל הם הפשוטים ביותר והם מנתחים, מקודדים ומשחזרים אות דיבור ברזולוציה של דגמים בודדים. מערכות עם דחיסה עמוקה יותר מנצלות תכונות זמניות כמו מחזוריות והשתנות איטית של עוצמת האות ו/או לחילופין תכונות של ספקטרום אות הדיבור. בקצבים נמוכים המערכות הן מסובכות יותר. מערכות אלו מניחות מודל מסוים של ייצור הדיבור. לרוב, אות דיבור מתואר כאות עירור העובר דרך מערכת לינארית משתנה בזמן, הממדלת את פונקצית התמסורת של חלל הפה. במקרה זה פרמטרי אות עירור וחלל הפה (מעטפת ספקטרלית) מקודדים בנפרד, דבר שמאפשר להוריד באופן משמעותי את קצב התמסורת של המקודד. במקודדי דיבור פרמטריים, אות הדיבור מחולק למסגרות (שאורכן 10-32ms) בהן המודל מניח סטציונריות של האות ולכל מסגרת כזו מחושבים פרמטרי המעטפת הספקטרלית ואות העירור. ההשהיה האלגוריתמית של מקודד דיבור כזה היא כגודל מסגרת האנליזה המקסימלית. מקודדים כאלו יכולים לתת איכות "סינטטית משופרת" (בין איכות סינטטית לאיכות תקשורת) בקצבים מעל 2 kbps. ע"מ להוריד את הקצב עוד יותר, יש לנצל תלות בין-מסגרתית

במקודדי דיבור פרמטריים, ע״ח הגדלת ההשהיה וסיבוכיות המערכת. עקב כך, מערכות אילו מתאימות יותר למערכות תקשורת half-duplex או למטרות אכסון של אותות הדיבור.

העבודה הנוכחית עוסקת בשיטות לייצוג יעיל של פרמטרי מקודדי דיבור מבוססי חיזוי לינארי תוך ניצול היתירות הזמנית שקיימת בין מסגרות דיבור עוקבות. עיקר תשומת הלב מופנה על טכניקות הקידוד המתבססות על מודל לפירוק הדיבור למאורעות זמניים או Temporal Decomposition (TD) [37]. TD היא שיטה למידול של רצף וקטורי הפרמטרים הספקטרליים ע״י אוסף מצומצם של מאורעות הדיבור המאופיינות ע״י וקטור פרמטרים יחיד ופונקצית אינטרפולציה, הממורכזת סביבו. השיטה החדשה לקידוד וקטורי פרמטרים ספקטרליים המוצעת בהמשך החיבור מתבססת על אלגוריתם ה-Optimal Restricted Temporal Decomposition (ORTD) [44] ומתאימה אותו לקידוד הקצבים הנמוכים ביותר (כ-300 סיביות לשנייה עבור המעטפת הספקטרלית) בזמן אמת. ביצועים של ה-ORTD המקורי השתפרו עקב התאמתו לקריטריון השגיאה הריבועית הממוצעת המשוקללת עם המשקלות הדינמיים התלויים בוקטורי הכניסה (Dynamically Weighted ORTD או DW-ORTD). בהמשך, האלגוריתם התת-אופטימלי פותח בהתבסס ב-ORTD, ע״מ להתאימו למימוש בזמן אמיתי. אלגוריתם זה, המכונה Sub-Optimal Restricted Temporal Decomposition (SORTeD) הוריד את העומס בחישובי של ה-ORTD בצורה משמעותית תוך כדי דגרדציה מזערית באיכות התאמת הפרמטרים הספקטרליים. שילוב של שני המודיפיקציות הללו הביא לאלגוריתם ה-DW-SORTeD (Dynamically Weighted SORTeD). אלגוריתם זה המשולב עם הקוונטיזציה של פרמטרי ה-TD, מאפשר קידוד מעטפת ספקטרלית של דיבור ב-300 bps עם השהיה אלגוריתמית של 160-250 ms במקודד. אלגוריתם זה שימש גם לדחיסת פרמטרי אות העירור (pitch ו-אנרגיה) לקבלת המקודד דיבור מלא לקצבים600-680 bps המבוסס על האנליזה MELP-2400 [6].

מבנה העבודה

העבודה מתמקדת כאמור בייצוג יעיל של פרמטרי הדיבור באמצעות מודל ה-Temporal Decomposition. פרק 2 נותן רקע למערכות קידוד דיבור, מוצג מבנה בסיסי של מערכת קידוד דיבור, האנליזה והסינתזה המקובלות במקודדי דיבור לקצבים נמוכים מסוג מקודדי חיזוי לינארי, ומדדי השגיאה להערכת איכות הקידוד של המעטפת הספקטרלית של הדיבור. פרק זה מכיל גם סקירה על מקודד העירור המעורב (MELP) שמהווה את הבסיס למערכת הקידוד לקצבים נמוכים מאד. פרק 3 מציג שיטות להורדת קצב השידור ובעיקר שתי גישות עיקריות: דילול מסגרות וקידוד משותף של מספר מסגרות. פרק 4 מתאר את מודל ה-Temporal Decomposition וסוקר יישומים מודרניים שלה לשימושי קידוד הדיבור, כולל ORTD. פרק 5 מציג אלגוריתם חדש לביצוע ה-TD עבור וקטורי המעטפת הספקטרלית (פרמטרי ה-LSF), המותאם ליחס דחיסה גבוה ושימוש בזמן אמת (DW-SORTeD). פרק 6 מתאר תוצאות ניסויים של טכניקת הקידוד שהוצגה בפרק 5. בפרק 7 מיישמים את אלגוריתם ה-DW-SORTeD לדחיסת פרמטרי עירור שונים במשותף (תדר ה-

pitch והאנרגיה). לבסוף, בפרק 8 מוצגת מערכת הקידוד המוצעת לקצבים של 600-650 bps, אשר מבוססת על התוצאות שהתקבלו בפרקים הקודמים. בדיקות מסכמות של המקודד מובאים גם הן בפרק 8. פרק 9 מסכם את העבודה ומציע כוונים להמשך המחקר.

פרק 2

רקע למערכת קידוד הדיבור

2.1 מבנה אות הדיבור

ע"מ ליצור מודל נאמן ליצירת הדיבור, יש להבין איך נוצר אות זה ומה הן תכונותיו העיקריות [1]. אות דיבור נוצר כתוצאה ממעבר האוויר הנדחף מהריאות דרך מיתרי הקול אל חלל הלוע, הפה והאף.

ניתן לסווג את אות הדיבור לשלושה סוגי צלילים עיקריים: צליל קולי (voiced), בו מיתרי הקול מהווים גורם עיקרי ביצירת אות והאות הוא בעל אופי מחזורי, צליל א-קולי (unvoiced), בו מיתרי הקול אינם רוטטים והאות הוא בעל אופי רועש, וצליל מעורב (mixed) המורכב משילוב של השניים הקודמים.

אותות קוליים נוצרים על ידי דחיפת אוויר מהריאות דרך שסע קטן בין מיתרי הקול (glottis). במצב ראשוני, מיתרי הקול נמצאים זה ליד זה , לחץ האוויר גורם לתנודות של מיתרי הקול, אשר נפתחים ונסגרים בצורה מחזורית, וכך מייצרים פולסי עירור מחזוריים (glottal pulses) שמעוררים את המעבר הקולי שמורכב מהגרון וחללי הלוע, האף והפה. ניתן לתאר את המעבר הקולי כאוסף חללי תהודה, אשר משתנים במהלך הדיבור עבור צלילים שונים. מחזור פולסי העירור נקרא המחזור היסודי או מחזור ה-pitch.

אותות א-קוליים נוצרים על ידי הצרות בנקודה מסוימת במעבר הקולי ודחיפת אוויר דרך נקודת ההצרות. פעולה זו יוצרת רעש רחב סרט המעורר את החלק שלאחר ההצרות במעבר הקולי. ניתן, בנוסף, להבחין בין צלילים א-קוליים בעלי עירור רעש בלבד, לבין צלילים א-קוליים פוצצים בהם יש עירורי הלם (לא מחזוריים) בנוסף לעירור רעש (כגון 'p', 'k', 'b').

אותות מעורבים נוצרים על ידי עירור המורכב משילוב של העירורים הקודמים, כלומר עירור מיתרי הקול ורעש (למשל 'v', 'z').

אותות קוליים הם כמעט מחזוריים בציר הזמן ובעלי מבנה כמעט הרמוני בציר התדר, לעומת זאת אותות א-קוליים הם דמויי רעש רחב סרט בציר הזמן. בנוסף לכך, בד"כ האנרגיה של קטע אות קולי גבוהה יותר מהאנרגיה של קטע אות א-קולי.

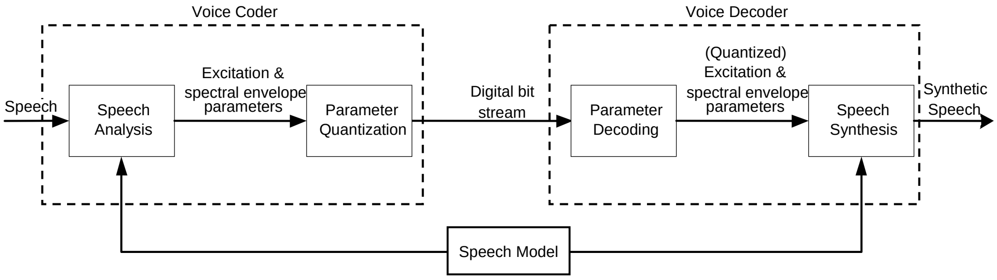
בכל סוגי הצלילים, פונקצית התמסורת של המעבר הקולי (vocal tract) אחראית ל"צביעת הספקטרום", כיוון שאות העירור (קולי וא-קולי) הנו בעל מבנה ספקטרלי בקרוב שטוח (רעש לבן או רכבת הלמים). באופן מעשי זה קובע איזה צליל קולי (phoneme) נהגה. תיאור מדויק של עצמת תגובת התדר של פונקצית תמסורת זו (המכונה גם "המעטפת הספקטרלית") בזמן סינתזה, מהווה

גורם מכריע להשגת מובנות גבוהה של הדיבור. לעומת זאת, תיאור מדויק של אות העירור (רעש ופולסים גלוטליים) תורם לשיפור איכות הצליל (במובן של טבעיות והעדר הפרעות).

המעטפת הספקטרלית של אות דיבור מאופיינת במספר שיאים הנקראים פורמנטים, אלו הם תדרי התהודה של המעבר הקולי. במערכת קולית ממוצעת ישנם עד ארבעה פורמנטים בתחום התדרים עד 4 kHz. האמפליטודות והמיקום של שלושת הפורמנטים הראשונים (שבדרך כלל נמצאים מתחת ל-3 kHz) חשובים בעת ביצוע סינטזת הדיבור מבחינת השמיעה האנושית.

2.2 מבנה מקודד דיבור פרמטרי

מערכות לקידוד דיבור בקצב נמוך (ר' איור 2-א) מורכבות משני מרכיבים עיקריים. האחד הוא אנליזה של אותות הדיבור לייצוג ע"י סט פרמטרי מודל, שנבחרו לאפיון אות הדיבור, והשני הוא כימות של פרמטרים אלה לסימבולים דיסקרטיים לשידור. המערכת הדואלית מפענחת את פרמטרי המודל המכומתים ומסנטזת בעזרתם את אות הדיבור



איור 2-א מערכת קידוד דיבור פרמטרי כללית

Figure 2-1 A generic parametric voice coder system

בשלב האנליזה מנתחים את אות הדיבור המוכפל בחלון סופי המחליק על פני האות על ציר הזמן עם גודל צעד קבוע (בד"כ) ומחשבים פרמטרי העירור והמעטפת הספקטרלית תוך הנחת סטציונריות של האות המשתקף מבעד חלון. בשלב הסינטזה מפיקים אות דיבור מתוך הפרמטרים המקוונטים בהסתמך על אותו מודל הדיבור.

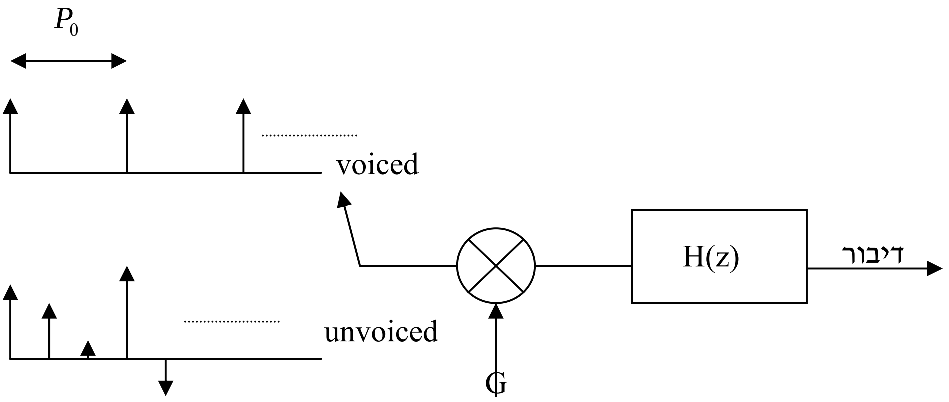
2.3 מודל ה-LPC של אות הדיבור ופרמטרי ייצוג של המעטפת הספקטרלית

2.3.1 המודל

המודל הבסיסי של אות דיבור קרוי מודל חיזוי לינארי (Linear prediction Coding /LPC/). השם מתייחס לדרך שערוך המעטפת הספקטרלית. מודל זה מניח עבור כל חלון אנליזה שהאות נוצר ע"י מעבר של עירור בעל ספקטרום שטוח דרך מערכת לינארית קבועה בזמן ויציבה בעלת L קטבים בלבד (הקרויה מסנן LPC):

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{i=1}^{p} a_i z^{-i}} \qquad (2.1)$$

מערכת זו ניתנת לאפיון חח"ע ע"י L פרמטרים (למשל L מקדמי הפולינום בפונקצית התמסורת, הקרויים מקדמי ה-LPC). אות עירור במודל זה יכול להיות מחזורי לייצוג צליל קולי או רעש לבן לייצוג צליל א-קולי (ר' איור 2-ב). עירור מחזורי פשוט ביותר מקורב ע"י רכבת הלמים (דגמי יחידה) בעלת מחזור בסיסי השווה למחזור ה-pitch של אותו קטע הדיבור. במקודדי דיבור בסיסיים המבוססים על חיזוי לינארי (LPC), המעטפת הספקטרלית מאופיינת ע"י מקדמי הפולינום של ה-LPC, ופרמטרי אות העירור - ע"י עוצמה, החלטה על סוג העירור (קולי/א-קולי) וערך מחזור ה-pitch. וקטור פרמטרים אלה משודר בד"כ 40-50 פעמים בשנייה [5]. מודל זה הינו פשוט ונוח בגלל מיעוט הפרמטרים הנחוצים לאפיון מסגרת דיבור וכן לנוכח קיום אלגוריתמים יעילים למציאת המעטפת הספקטרלית [2].



איור 2-ב. מערכת קידוד דיבור LPC הבסיסית.

Figure 2-2. The basic LPC Model

מעטפת ה-LPC מאפשרת ליצור אות דיבור בעל מובנות גבוהה ע" מספר קטן יחסית של פרמטרים. ייצוג אות עירור במודל הבסיסי הוא לא עשיר מספיק מכיוון שהוא לא נותן ייצוג נאות לצלילים מעורבים (שילוב בין קולי וא-קולי, מעבר בין קולי לא-קולי) או מסגרות אשר לא מתאימות למודל (למשל אם בפועל קיים אפס בפונקצית התמסורת). כמו כן מודל זה רגיש מאד לשגיאות ההחלטה על סוג העירור. עקב הבעיות הללו הדיבור שמופק ע"י מקודד זה נשמע מזמזם ו"מתכתי". אי לכך פותחו מספר רב של מקודדי דיבור מבוססי LPC, אך בעלי ייצוג אות עירור עשיר יותר [8,6]. במקודדים אלו העירור בכל מסגרת הוא שילוב של מרכיב מחזורי (לאו דווקא בעל ספקטרום שטוח) ומרכיב רועש. העירוב בין שני מרכיבים אלו יכול להיות שונה בפסי תדר שונים. מקודדים אלו מאפשרים שיפור איכות הצליל מבלי לפגוע ביעילות ונוחות הייצוג ע"י מעטפת ה-LPC. כאמור, ייצוג

יעיל של מעטפת ספקטרלית יכול להביא למובנות גבוהה של אות דיבור גם בקצבים נמוכים, אך יצירת דיבור בעל איכות הקרובה למקור מחייבת הגדלת הקצב המוקדש לתיאור אות העירור. לדוגמא, בתקן MELP (לקצב 2400 bps), כ-50% מהסיביות מיועדות לתיאור אות העירור [6], לעומת תקן המממש מודל בסיסי לאותו קצב (LPC-10) שמקצה כ-20% בלבד לייצוג אות העירור [5].

2.3.2 שערוך המעטפת הספקטרלית באמצעות חיזוי לינארי

שערוך לינארי הוא הדרך הנפוצה מיותר למציאה ופרמטריזציה של המעטפת הספקטרלית של הדיבור. היא מאפשרת לשערך את הספקטרום של אות הדיבור ללא צורך בקביעת סוג העירור (קולי/א-קולי) ובצורה יעילה חישובית.

החזאי הלינארי מוגדר לפי:

$$\hat{s}(n) = \sum_{i=1}^{p} a_i s(n-i) \tag{2.2}$$

כאשר, $s(n)$ הוא אות הדיבור ו-$\hat{s}(n)$ הוא האות שמשוערך על פי p הדגימות האחרונות. מקובל למצוא את מקדמי החזאי האופטימלי על ידי ביצוע מינימיזציה של שגיאת השערוך (נקרא גם אות השארית) על פני קטע זמן נתון:

$$e(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \tag{2.3}$$

כאשר ממזערים את:

$$\varepsilon^2 = \sum_{n=n_0}^{n_1} e^2(n) \tag{2.4}$$

n_0 ו- n_1 מגדירים את תחום המינימיזציה. בשיטת האוטוקורלציה [1] תחום המינימיזציה מוגדר על פני כל ציר הזמן. מכיוון שמספר הדגימות הוא סופי (N), עקב ההכפלה בחלון, הרי שתחום המינימיזציה הוא למעשה $[0, N-1+p]$. אם נגדיר את $r(k)$:

$$r(k) = \sum_{n=0}^{N-1-k} s(n)s(n+k) \tag{2.5}$$

אזי המשוואות המתקבלות מגזירת הביטוי (2.4) לפי המקדמים $\{a_i\}_{i=1}^{p}$ והשוואה לאפס הן:

$$\sum_{i=1}^{p} a_i r(i-k) = r(k), \quad k = 1,2,...,p \tag{2.6}$$

או ברישום מטריצי:

$$\mathbf{Ra} = \mathbf{r} \tag{2.7}$$

כאשר, **R** היא מטריצת טופליץ (Toeplitz) מסדר $p \times p$:

$$\mathbf{R} = \begin{pmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & r(1) \\ r(p-1) & r(p-2) & \cdots & r(0) \end{pmatrix}. \tag{2.8}$$

ניתן לפתור את מערכת המשוואות (2.7) בצורה יעילה למציאת מקדמי החזאי הלינארי האופטימלי a בעזרת אלגוריתם Levinson-Durbin [1]. שיטה זו מניחה התאפסות של האות מחוץ לחלון האנליזה, לכן ע״מ להקטין שגיאת השערוך עקב תנאי קצה, משתמשים בחלון אנליזה ארוך יחסית (כ-30 מילישניות) המתקרב לאפס בקצוות (למשל חלון Hamming). קיימת שיטה נוספת לשערוך הספקטרום,המכונה שיטת ה-covariance [1], אשר היא מדויקת יותר, מכיוון שאינה מניחה דבר לגבי האות מחוץ לחלון האנליזה. שיטה זו כבדה יותר מבחינה חישובית, אך מאפשרת לעשות אנליזה מדויקת יותר בחלונות קטנים יותר (חלון הקטן מאורך מחזור-pitch בודד). ע״מ להקטין את הרגישות של שיטה זו במיקום היחסי של חלון האנליזה ותחילת מחזור ה-pitch (נקודת הסגירה של מיתרי הקול המייצרת שיא דומיננטי בקרבתה, המכונה Glottal Closure Instant (GCI), יש לבחור חלונות בצורה סינכרונית ל-GCI. עובדה זו מתנה דיוק של שערוך בשיטת ה-covariance במספר החלטות אמפיריות הקשורות במציאת מחזור ה-pitch ונקודת ה-GCI.

2.3.3 ייצוג מסנן ה-LPC בעזרת LSF

קיים מגוון רחב של פרמטרים לייצוג מסנן ה-LPC שמספרם תמיד כסדר המסנן, כגון : קטבי המסנן, log-areas (LA), log-area ratios (LAR), מקדמי החזרה (reflection coefficients), line spectral frequencies (LSF) המכונים גם לעתים line spectral pairs (LSP) ועוד. הפרמטרים שיעילים במיוחד לקוונטיזציה הם מקדמי ה-LSF. מקדמי ה-LSF הוצגו לראשונה ב-[54] ומסתבר שהם בעלי רגישות נמוכה יותר לקוונטיזציה (סקלרית ואף וקטורית) מפרמטרי ייצוג אחרים [9] וכן מתאימים במיוחד לאינטרפולציה לינארית לייצוג ספקטרום במסגרות חסרות [10]. ניתן למצוא פרמטרים אלה בדרך הבאה: נגדיר את החזאי הלינארי האופטימלי מסדר p בעזרת הפרמטרים $\{a_i\}_{i=1}^{p}$. ניתן להראות [1] כי מתקיים הקשר הרקורסיבי הבא (מסנן lattice) :

$$A_j(z) = A_{j-1}(z) - k_j z^{-j} A_{j-1}(z^{-1}), \quad j = 1,2,...,p \tag{2.9}$$

כאשר, הפרמטרים $\{k_i\}_{i=1}^{p}$ נקראים מקדמי החזרה (Reflection Coefficients או PARCOR) ו-$A_j(z)$ היא פונקצית התמסורת של מסנן הסינתזה מסדר j. אם נוסיף למסנן ה-lattice חוליה $p+1$ עם תנאי גבול מלאכותיים $k_{p+1}=1$ ו- $k_{p+1}=-1$ המתאימים לסגירה ופתיחה מוחלטים של ה-glottis, נקבל את הפולינומים הבאים :

$$
\begin{aligned}
P_{p+1}(z) &= A_p(z) + z^{-(p+1)}A_p(z^{-1}) \\
Q_{p+1}(z) &= A_p(z) - z^{-(p+1)}A_p(z^{-1})
\end{aligned}
\qquad (2.10)
$$

הפולינומים $P(z)$ ו-$Q(z)$ הם בעלי תכונות מעניינות [54], השורשים שלהם נמצאים על מעגל היחידה והשורשים של שני הפולינומים שזורים אלו באלו (תנאי הכרחי ומספיק ליציבות המסנן, הקרוי *תכונת הסדר* (*ordering property*) של ה-Line Spectral Frequencies). אפשר לייצג את הפולינומים מ-(2.10) באופן הבא (עבור חזאי מסדר זוגי) :

$$
\begin{aligned}
P_{p+1}(z) &= (1+z^{-1}) \prod_{i=1,3,5,\ldots,p-1} (1-2\cos\omega_i z^{-1} + z^{-2}) \\
Q_{p+1}(z) &= (1-z^{-1}) \prod_{i=2,4,\ldots,p} (1-2\cos\omega_i z^{-1} + z^{-2})
\end{aligned}
\qquad (2.11)
$$

סט הפרמטרים $\{\omega_i\}_{i=1}^{p}$ נקראים Line Spectral Frequencies (LSF). מתכונת הסדר של ה-LSF נובע, כי $\omega_1 < \omega_2 < \ldots < \omega_p$. אם ערכי ה-LSF השכנים קרובים זה לזה, הדבר מעיד על הימצאות פורמנט בקרבתם. ע"י הגבלת המרחק המינימלי המותר בין ערכי ה-LSF השכנים ניתן למנוע שיאים חדים בספקטרום אשר עלולים להביא לאי-יציבות נומרית. מקדמי ה-LSF משתנים לאט ממסגרת למסגרת ומכיוון שערכיהם קשורים למיקום הפורמנטים בספקטרום, אינטרפולציה לינארית בין וקטורי ה-LSF תייצג בד"כ ספקטרום ביניים בו הפורמנטים "נעים" ממיקומם בווקטור ה-LSF הראשון למשנהו. יתרה מזאת, אינטרפולציה לינארית בין מקדמי ה-LSF המופקים ממסנני LPC יציבים תביא בהכרח לייצוג LSF של מסנן יציב. עקב תכונת המקומיות הספקטרלית של מקדמים אלו, רעש קוונטיזציה של מקדם LSF מסוים משפיע בעיקר על הספקטרום בקרבתו, ולא על המעטפת הספקטרלית כולה. ניתן לנצל זאת ע"י הקצאה שונה של סיביות בהתאם לתכונות השמיעה האנושית.

2.4 פונקצית עוות של מקדמי מסנן ה-LPC

בסעיף זה יוצגו פונקציות עוות אשר משמשות בקוונטיזציה ווקטורית של מקדמי מסנן ה- LPC או התמרה מקובלת שלהם למקדמי Line Spectral Frequencies – LSF.

2.4.1 מדד ה-Log-Spectral Distortion (LSD)

אחד המדדים המקובלים למדידת הקרבה הספקטרלית (המותאמת לשמיעה אנושית) הוא עוות מסוג Log Spectral Distortion (LSD ביחידות של dB). העיוות נמדד בין וקטור ה-LPC לפני הקוונטיזציה ואחריה. המדד מוגדר לפי:

$$d_{LSD}(A,\hat{A}) = \sqrt{\frac{1}{2\pi}\int_{-\theta_1}^{\theta_2}\left(10\log_{10}\frac{1}{|A(\omega)|^2} - 10\log_{10}\frac{1}{|\hat{A}(\omega)|^2}\right)^2 d\omega} =$$

$$= \sqrt{\frac{\beta}{2\pi}\int_{-\theta_1}^{\theta_2}\left(\ln|A(\omega)|^2 - \ln|\hat{A}(\omega)|^2\right)^2 d\omega} \qquad (2.12)$$

$$\beta = (10/\ln 10)^2, \qquad A(\omega) = 1 - \sum_{i=1}^{v} a_i e^{j\omega i}, \qquad \hat{A}(\omega) = 1 - \sum_{i=1}^{v} \hat{a}_i e^{j\omega i}$$

כאשר θ_2, θ_1 הם הגבולות רוחב הסרט האפקטיבי של הערוץ בו מועבר הדיבור.

העיוות נמדד במישור הלוגריתמי עקב תכונת הרגישות הלוגריתמית של האוזן האנושית. יחד עם המדד הממוצע של LSD על פני קטעי דיבור נבדקים, מקובל לבחון גם אחוז המסגרות החריגות בהרבה מהממוצע. ב-[13] מוגדרת איכות קוונטיזציה "שקופה" כפי שמצוין בטבלה 2-א.

Table 2-a. Transparent quality of spectral envelope as defined by the LSD measure.

טבלה 2-א. איכות שקופה של מעטפת ספקטרלית של הדיבור, כפי שמוגדר ע"י מדד ה-LSD.

Average LSD, [dB]	Frames with LSD in [2..4] dB, [%]	Frames with LSD greater then 4dB, [%]
≤ 1	<2	0

2.4.2 מדדים המבוססים על שגיאה ריבועית משוקללת

למרות שמדד עוות זה טוב ומקובל לבדיקת ביצועי מקוונטים (quantizers) שונים, קיימת בעיה של סיבוכיות תכן מקוונטים למזעור שגיאה זו. לכן, לרוב מתכננים מקוונטים למזעור מדד עיוות פשוט יותר כגון עוות ריבועי (MSE) או עוות ריבועי משוקלל (WMSE):

$$d_{MSE} = (\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}}), \qquad (2.13)$$
$$d_{WMSE} = (\mathbf{x}-\hat{\mathbf{x}})^T \mathbf{W}_{\mathbf{x}} (\mathbf{x}-\hat{\mathbf{x}})$$

כאשר $\mathbf{W_x}$ היא מטריצת משקלות אלכסונית שיתכן והיא תלויה בווקטור הכניסה $\mathbf{x}$.

דוגמא לכך ניתן לראות בסכמה שהוצעה על ידי Paliwal ו-Atal [13] לתכן מכמת וקטורי למקדמי ה-LSF עם 24 סיביות. פונקצית העיוות היא:

$$d_{WMSE} = \sum_{i=1}^{10} [c_i w_i (\omega_i - \hat{\omega}_i)]^2 \tag{2.14}$$

כאן ω_i ו- $\hat{\omega}_i$ הם איברי וקטורי ה-LSF לפני ואחרי הכימות, בהתאמה. w_i היא המשקל התלוי בווקטור הכניסה אשר מחושב לפי:

$$w_i = [P(\omega_i)]^r \tag{2.15}$$

כאשר, $r = 0.15$ נקבע באופן אמפירי ו- $P(\omega)$ הוא ספקטרום ההספק של מסנן הסינתזה :

$$P(\omega) = \frac{1}{|A(\omega)|^2} \tag{2.16}$$

c_i ב-(2.14) הנו פקטור קבוע נוסף המפחית את חשיבות של התדרים הגבוהים במדד העיוות הכולל:

$$c_i = \begin{cases} 1.0 & 1 \le i \le 8 \\ 0.8 & i = 9 \\ 0.4 & i = 10 \end{cases} \tag{2.17}$$

אנו נקרא מכאן ואילך למדד זה בשם שגיאה ריבועית ממוצעות משוקללת של Paliwal ו-Atal, או PA-WMSE. יש לציין כי עוות זה הוא אמפירי, ולכן אין קשר ברור בינו לבין עוות ה-LSD.

קריטריון שגיאה ריבועית משוקללת נוסף פותח על ידי Gardner [550]. הוא מראה בעבודתו כי ניתן לקרב את עוות ה-LSD על ידי עוות WMSE עם מטריצת משקלים מתאימה. בצורה כזו, שימוש בעיוות ה-WMSE החדש יוביל לתוצאות טובות יותר של ה-LSD (עבור קוונטיזציה צפופה מספיק).

קירוב ה-LSD ע"י WMSE ניתן למצוא ע"י הסתכלות על פירוק טיילור (Tailor) של פונקצית עוות כללית (חלקה מספיק). נרשום את טור טיילור של פונקצית עוות זו, $d(\mathbf{x},\hat{\mathbf{x}})$, סביב $\mathbf{x} = \hat{\mathbf{x}}$:

$$d(\mathbf{x},\hat{\mathbf{x}}) = d(\hat{\mathbf{x}},\hat{\mathbf{x}}) + D(\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^T \mathbf{W}(\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}}) + O(\|\mathbf{x}-\hat{\mathbf{x}}\|^3) \tag{2.18}$$

כאשר, $\mathbf{D}(\hat{\mathbf{x}})$ הוא וקטור הגרדיינט של פונקצית העיוות שרכיביו ניתנים ע"י:

$$D_j(\hat{\mathbf{x}}) = \left.\frac{\partial d(\mathbf{x},\hat{\mathbf{x}})}{\partial x_j}\right|_{\mathbf{x}=\hat{\mathbf{x}}} \tag{2.19}$$

ו- $\mathbf{W}(\hat{\mathbf{x}})$ היא מטריצת Hessian, שאיבריה ניתנים ע"י:

$$W_{j,k}(\hat{\mathbf{x}}) = \left.\frac{\partial^2 d(\mathbf{x},\hat{\mathbf{x}})}{\partial x_j \partial x_k}\right|_{\mathbf{x}=\hat{\mathbf{x}}} \tag{2.20}$$

מההגדרה של פונקצית העיוות מקבלים שהאבר הראשון $d(\hat{\mathbf{x}},\hat{\mathbf{x}})=0$. כיוון ש-$\hat{\mathbf{x}}$ הנו מינימום מקומי של פונקצית העיוות $d(\mathbf{x},\hat{\mathbf{x}})$, מתקיים גם, כי $D(\hat{\mathbf{x}})=\mathbf{0}$ ולכן:

$$d(\mathbf{x},\hat{\mathbf{x}}) = \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^T \mathbf{W}(\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}}) + O(\|\mathbf{x}-\hat{\mathbf{x}}\|^3) \tag{2.21}$$

הדרישה ש- $d(\mathbf{x},\hat{\mathbf{x}}) \geq 0$ גוררת ש-$\mathbf{W}(\hat{\mathbf{x}})$ היא מטריצה positive semi-definite.

עבור ספר קוד צפוף ניתן להזניח את השארית $O(\|\mathbf{x}-\hat{\mathbf{x}}\|^3)$, ולכן:

$$d(\mathbf{x},\hat{\mathbf{x}}) = \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^{\mathbf{T}} \mathbf{W}(\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}}) \tag{2.22}$$

Gardner מראה, בנוסף, כי בהנחת ספרי קוד צפופים ניתן להשתמש בקירוב הבא לפונקצית העיוות:

$$\tilde{d}(\mathbf{x},\hat{\mathbf{x}}) = \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^T \mathbf{W}(\mathbf{x})(\mathbf{x}-\hat{\mathbf{x}}), \quad \hat{\mathbf{x}} \triangleq Q(\mathbf{x}), \tag{2.23}$$

כלומר ניתן להשתמש במטריצת משקלות אשר תלויה בוקטורי הכניסה בלבד – דבר שמאפשר להשתמש בקירוב זה ל-LSD לצורך אימון ספר קוד וקידוד באמצעותו.

באופן כללי, המטריצה $\mathbf{W}(\mathbf{x})$ איננה אלכסונית, אך במקרה של פרמטרי ה-LSF המטריצה היא אכן אלכסונית והיא ניתנת לחישוב ע"י:

$$\mathbf{W} = 4\beta \mathbf{J}^T \mathbf{R}_A \mathbf{J} \tag{2.24}$$

כאשר $\mathbf{R}_A$ היא מטריצת האוטוקורלציה של התגובה להלם $\{h(n)\}$ של $\frac{1}{A(z)}$, β הוא קבוע לפי (2.12) ו- $\mathbf{J}$ היא מטריצת היעקוביאן של הטרנספורמציה ממקדמי ה-LPC $\{a_n\}_{n=1}^{p}$ למקדמי ה-LSF $\{\omega_i\}_{i-1}^{p}$, שאיבריה ניתנים ע"י:

$$J_{i,k} = \left.\frac{\partial a_k(\hat{\boldsymbol{\omega}})}{\partial \omega_i}\right|_{\hat{\boldsymbol{\omega}}=\boldsymbol{\omega}} . \tag{2.25}$$

אנו נכנה מכאן ואילך מדד שגיאה זו - שגיאה ריבועית ממוצעת משוקללת של Gardner, או G-WMSE.

2.5 עירור מעורב המשולב עם מערכת חיזוי לינארי (MELP)

הזכרנו קודם, כי העשרה של מודל העירור היא המפתח לשיפור באיכות הדיבור המסונטז במערכות מבוססות חיזוי לינארי. מקודד הידוע בשם Mixed Excitation Linear Prediction או MELP [6],

הוא אחת השיטות להעשרת מודל העירור. הוספת הפרמטרים הנוספים (בנוסף לאילו שהזכרנו עבור מודל ה-LPC הבסיסי, כגון אנרגיה, תדר ה-pitch וסווג קולי/א-קולי) נועדים לשיפור ההתאמה הספקטרלית של האות המסונטז למקור וגם ייצוג נאמן יותר של העירור המעורב (רעש יחד עם האות עירור מחזורי).

בסעיף זה נתאר בקווים כלליים את פרמטרי העירור של מקודד העירור המעורב,MELP [6]. מקודד זה משפר את אות הדיבור המשוחזר שמתקבל במקודד ה-LPC המבוסס על סינתזה של אות קולי או א-קולי בלבד (החלטת V/UV אחת). השיפורים העיקריים של העירור ב-MELP, לעומת מקודד LPC פשוט [5], כוללים עירור מעורב של פולסים ורעש לבן בחמישה פסי תדר שונים, אפשרות של שימוש בפולסים לא מחזוריים עבור דיבור קולי, וקידוד חלקי של ספקטרום אות העירור. להלן הפרמטרי אות העירור שמשודרים כל מסגרת (22.5 מילישניות):

1) <u>ערך ה-pitch.</u> אם ערכו שווה לאפס הדבר מסמן כי המסגרת היא א-קולית, כלומר כל חמשת פסי התדר הם א-קוליים, אם לאו, פס התדר הראשון (0-500 Hz) הוא קולי, ופרמטרי ה-voicing של ארבעת פסי תדר הבאים יקבעו מידת הקוליות (ה-voicing) של המסגרת. עקב קיום פרמטרי ה-voicing האלה ופרמטרים נוספים (כגון jitter), שיכולים לפצות על קביעת pitch השונה מאפס במקומות הלא נכונים, האלגוריתם לקביעת ה-pitch נוטה להחליט על קיום ה-pitch כמעט תמיד, למעט מסגרות שהן "א-קוליות בוודאות".

2) <u>האנרגיה.</u> נמדדת פעמיים למסגרת. (גם קידוד דיפרנציאלי של השנייה לעומת הראשונה).

3) <u>החלטות הקוליות (מסגרות קוליות בלבד).</u> החלטות בינריות לגבי הקוליות של ארבעה פסי התדר (Hz): 500-1000, 1000-2000, 2000-3000, 3000-4000. (החלטת הקוליות על הפס הראשון (0-500 Hz) קובעת למעשה קוליות של המסגרת כולה.)

4) <u>ספקטרום חלקי של פולס העירור (מסגרות קוליות בלבד).</u> עשרה מקדמים ראשוניים של התמרת התדר של פולס העירור משודרים (דחוסים בעזרת VQ). מקדמים אלה מציינים, למעשה, עד כמה שונה ספקטרום העירור מספקטרום לבן (שטוח).

5) <u>Jitter או Aperiodicity flag (מסגרות קוליות בלבד).</u> דגל בינרי זה מציין האם יש להרעיד את מיקומי פולסי העירור, במידה והמסגרת אינה "מחזורית מספיק" ע"מ להימנע מזמזומים לא רצויים בצלילי מעבר בסינטזה.

הקצאת הסיביות למסגרת במקודד MELP-2400 מוצגת בטבלה 2-ב.

Table 2-b. MELP coder bit allocation for a single frame

טבלה 2-ב. הקצאת סיביות למקודד דיבור MELP עבור מסגרת בודדת

Parameter/פרמטר	**Voiced**	**Unvoiced**
LSF (10 parameters)	25	25
Fourier Magnitude (10 parameters)	8	-
Gain (2 per Frame)	8	8
Pitch , overall voicing	7	7
Bandpass voicing	4	-
Aperiodic Flag (Jitter)	1	-
Error Protection	-	13
Sync bit	1	1
Total bits / 22.5ms frame	54	54

ניתן לראות, כי המעטפת הספקטרלית המיוצגת בעזרת LSF מהווה כ- 45% מסך כל הסיביות למסגרת דיבור בודדת. בפרקים הבאים יוצגו טכניקות להורדת קצב השידור שמבוססות בעיקר על הורדת קצב השידור של המעטפת הספקטרלית. בפרק 3 יוצגו מספר שיטות להורדת קצב שדווחו בספרות ובפרק 5 יוצג אלגוריתם חדש המבוסס על מודל ה-Temporal Decomposition (שמוסבר בפרק 4). בנוסף, בפרק 6 תוצע דרך לדחיסה עמוקה של פרמטרי העירור העיקריים (תדר ה-pitch והאנרגיה).

פרק 3

שיטות לקידוד פרמטרים ספקטרליים

3.1 מבוא

בפרק זה נסקור מספר שיטות ידועות לכימות פרמטרים ספקטרליים במקודדי דיבור מבוססי LPC בקצב נמוך מאד. שיטות אלו מנצלות תלות זמנית בין פרמטרים ספקטרליים ע״מ לצמצם את קצב השידוד על חשבון הוספת השהיה למערכת, הגדלת הסיבוכיות ואף התרת קצב שידור משתנה.

באחרונה הושקע הרבה מאמץ מחקרי בפיתוח מקודדי דיבור לקצבים נמוכים מאוד (מתחת ל-2000 סיביות לשנייה), כולל הגדרת תקן חדש, מבוסס MELP לקצב של 1200 סיביות לשנייה [7].

3.2 קידוד פרמטרים ספקטרליים

שיטות כימות רבות פותחו לייצוג פרמטרים ספקטרליים בכמות קטנה ככל האפשר של סיביות. כימות סקלרי מייצג את כל אחד מהפרמטרים הספקטרליים בנפרד. כימות סקלרי הינו פעולה פשוטה יחסית אך ביצועיה מוגבלים. כימות סקלרי של הפרמטרים בכל מסגרת מביא לקצב כולל של כ-2400 סיביות בשנייה במקודד LPC בעל איכות קול סינטטית (הפרמטרים הספקטרליים משודרים בקצב של כ-40 סיביות למסגרת) [5].

בכימות וקטורי (VQ) מייצגים אוסף של פרמטרים ספקטרליים כווקטור אחד [16,14,12]. ספר קוד מכיל אוסף וקטורים המייצגים כ״א פרמטרים המתאימים למסגרת זמנית אחת. כימות וקטורי מנצל בצורה יעילה תלות תוך-מסגרתית בין הפרמטרים השונים. ע״י שימוש בקידוד וקטורי (עם חיפוש מלא) , ניתן לקבל קידוד ספקטרלי באיכות טובה בקצב של 20 סיביות למסגרת עבור הפרמטרים הספקטרליים [14]. החסרונות של שיטת VQ מלא הינם דרישות זיכרון וכוח חישוב גבוהים. כמו כן, גודל סדרות האימון הנדרש למציאת ספר קוד טוב הינו עצום.

ישנן שיטות של כימות וקטורי אשר משתמשות בספרי קוד או אלגוריתמי חיפוש תת-אופטימליים [15,02], או באלו שמפצלות את הפרמטרים הספקטרליים למספר וקטורים המקודדים בנפרד (Split VQ) [13], ומביאות לחסכון ניכר בעומס החישובי ו/או דרישות הזיכרון. בשיטות אילו מקבלים ביצועים דומים לאלו של VQ ישיר אך במחיר הגדלת קצב שידור הפרמטרים הספקטרליים ב-3-4 סיביות למסגרת [13].

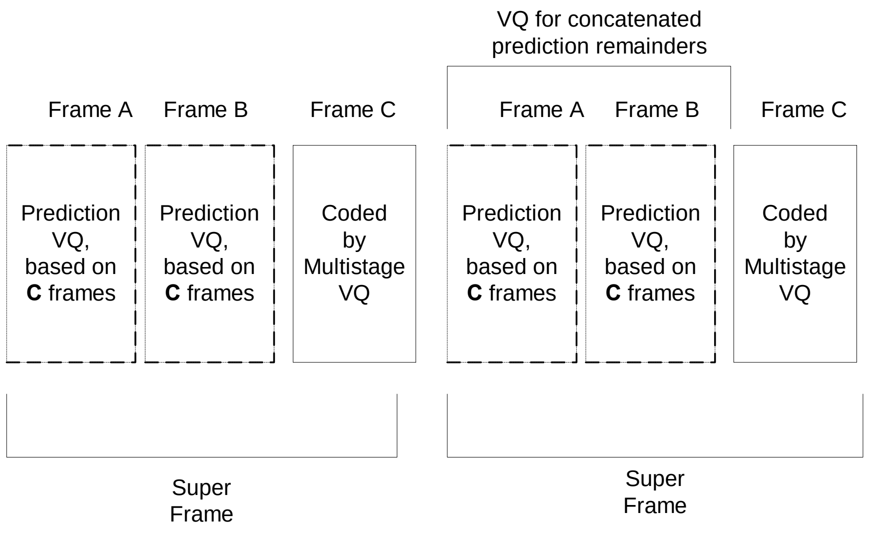
ע״מ להוריד עוד את קצב העברת הפרמטרים הספקטרליים (וכתוצאה מכך את הקצב כולל של המקודד), יש לנצל גם את התלות הזמנית (הבין-מסגרתית) של פרמטרי הספקטרום, דבר שעלול

לדרוש השהיות ארוכות יחסית ויקשה על שימוש מקודדים כאלה בשימושים רבים. מכיוון שפנינו לשימושים המאפשרים השהיה ארוכה יחסית (מעל 200 ms), יש בדעתנו לנצל במידה רבה את התלות הזמנית.

3.3 ניצול תלות זמנית עם השהיה קצרה

בין שיטות הכימות הווקטורי הקיימות ישנן שיטות תת-אופטימליות אשר נועדו לשימושים הדורשים השהיה קצרה, כמו כימות וקטורי עם חיזוי לינארי (PVQ) [21,20,19] או כימות וקטורי בשילוב עם מכונת מצבים סופית (FSVQ) [17,18].

בשיטה הראשונה (PVQ - Prediction VQ) מכמתים וקטור הפרש בין וקטור הכניסה לבין מוצא חזאי לינארי אופטימלי לאותו וקטור (סדר החזאי הווקטורי הוא בד"כ אחד או שניים). ברעיון זה משתמשים בקידוד ה-LSF בתקן MELP-1200 החדש [7]. בתקן זה מקודדים כל פעם בלוק של 3 מסגרות קוליות עוקבות, כאשר המסגרת השלישית (נסמנה C) מקודדת ע"י VQ רב-שלבי (קידוד אבסולוטי), והשאר באמצעות החיזוי: תחילה, הערכים החזויים של ווקטורים אלו מחושבים כאינטרפולציה לינארית של זוג וקטורי C (מהשלישייה הנוכחית והשלישייה הקודמת), כאשר מקדמי האינטרפולציה נבחרים מתוך ספר קוד מצומצם (בעל 4 סיביות), ולאחר מכן שגיאות החיזוי של 2 וקטרים אלו (משורשרות יחד) מקודדות באמצעות VQ רב-שלבי נוסף (ר' איור 3-א). העובדה, כי פרמטרים של שתי מסגרות מקודדות יחד מזכירה גם גישת הקוונטיזציה המטריצית (MQ) אשר תידון בסעיף 3.5.



איור 3-א. PVQ קדמי-אחורי לקידוד וקטורי ה-LSF בתקן MELP-1200.

Figure 3-1. Forward-Backward PVQ, as used in MELP-1200 standard for LSF quantization

בשיטה השנייה (FSVQ) משתמשים באוסף של מכמתים ווקטוריים חסרי זיכרון, כאשר בחירת המכמת נקבעת בהתאם למצב במכונת מצבים סופית.

החסרונות של השיטות הללו הינם רגישות לשגיאות ערוץ (תופעת התפשטות השגיאה) ותת-אופטימליות, בהשוואה לשיטות אחרות כמו כימות מטריצי וקידוד סגמנטים (ר' בהמשך). ב-[190] משלבים בין השיטות הנ"ל (FSVQ , PVQ) וספר קוד קבוע (המשמש כ"רשת ביטחון") למניעת התפשטות השגיאה בערוץ רועש ומדווחים שם על שיפור של כ-4 סיביות למסגרת בהשוואה לשיטת Split-VQ.

3.4 שיטות המתבססות על דילוג מסגרות

קיים מגוון רחב של שיטות (הן בקצב קבוע, הן בקצב משתנה) בהם לא משדרים את מלוא המידע בכל מסגרת זמנית. המידע הספקטרלי החסר משוחזר במפענח ע"י אינטרפולציה. למשל ב-[26] משדרים פרמטרים ספקטרליים כל מסגרת שנייה וב-[27] בוחרים מכל בלוק בן 8 מסגרות 4 מסגרות המביאים למינימום שגיאת בלוק כוללת, כאשר בשאר מסגרות הבלוק משערכים את פרמטרי הספקטרום ע"י אינטרפולציה. השיטה המוצעת ב-[30], הנקראת TSQ (Trellis Segment Quantization), משלבת את הבחירה הנ"ל עם אפשרות כימות של מספר מסגרות עוקבות במשותף. מוצע שם אלגוריתם סבכה (trellis) יעיל לבחירת הוקטורים המיצגים, אשר בהינתן ספר הקוד, ממזער את השגיאה הריבועית המשוקללת. (שיטה זו, המשלבת כימות סגמנטים ודילוגים על מסגרות גם יחד, תיסקר בהרחבה בסעיף 3.6 בהמשך). ב-[28,27] מציעים אלגוריתם לקידוד פרמטרים ספקטרליים בקצב משתנה (עם אילוץ של השהיה שלא עולה על 200 ms). בשיטה זו, הנקראת CQI (Combined Quantization Interpolation), משדרים רק את הוקטורים המבטאים שינויים ספקטרליים חדים, ואת וקטורי פרמטרים ושאר המסגרות משערכים במפענח ע"י אינטרפולציה. בחירת המסגרות המייצגות נעשית בשיטה תת-אופטימלית פשוטה, המורכבת משני מעברים על הנתונים. במעבר הראשון בוחרים את הנציגים המרוחקים אחד מהשני ככל הניתן, כל עוד השגיאה הריבועית המשוקללת אינה עולה על סף נבחר. במעבר השני, מוצאים מיקום אופטימלי מבחינת מינימום שגיאה של כל נציג ביחס ל-2 הנציגים הסובבים אותו. שיטה זו נבחנה בקצבים ממוצעים של כ-300 סיביות לשנייה והניבה LSD ממוצע של 2.84 ו- 3.07 dB עבור קצבים ממוצעים של 320 ו-256 סיביות לשנייה בהתאמה. שיטה זו מנצלת תלויות לינאריות בלבד בין מסגרות עוקבות וכן תלויה בפרמטרים אמפיריים.

3.5 כימות מטריצי (MQ)

הרחבה ישירה של VQ היא שיטת הכימות המטריצי (Matrix Quantization או MQ), בה מתבצע כימות וקטורי של מספר קבוע של מסגרות זמניות במשותף. תהי Y מטריצת פרמטרים ספקטרליים

בגודל $P \times N$ המכילה וקטורי פרמטרים ספקטרליים עוקבים, במימד p כל אחד, כוקטורי עמודה. (לכל מסגרת זמנית מופקים p פרמטרים המתארים את המעטפת הספקטרלית של אות הדיבור במסגרת):

$$\mathbf{Y} = \left[\mathbf{y}(1) \quad \mathbf{y}(2) \quad \cdots \quad \mathbf{y}(N)\right]_{p\times N} \tag{3.1}$$

הרעיון הוא לייצג N וקטורי פרמטרים ספקטרליים עוקבים (Y) בעזרת מילון הכולל 2^K מילות קוד (שהן מטריצות במימד $p \times N$). Tsao ו-Gray מציינים ב-[22] שככל שמגדילים את מימד המטריצה (N), ומייצרים ספר הקוד תוך שמירה על עוות נתון, קצב הסיביות למסגרת K/N הולך וקטן (K גדל לאט יותר מאשר N לעוות קבוע), ולכן ניתן להקטין את קצב השידור במחיר הגדלת סיבוכיות ודרישות זיכרון. קיים עדות לכך [22] שאין שיפור משמעותי בשבח הקידוד עבור $N > 4$. ז״א, קידוד משותף של 4 מסגרות יחד ממצה את רוב התלויות הבין-מיסגרתיות, דבר שמתיישב עם קצב הפונמות בדיבור (כ-12 בשנייה).

הבעיה העיקרית בשיטת MQ הינה העומס החישובי ודרישות הזיכרון הגבוהים [23]. קיימות מספר שיטות מבוססות MQ, שחותרות להקטין את הדרישות הללו. ב-[24] מפרקים את מטריצת הפרמטרים הספקטרליים Y למטריצת הצנטרואידים S ומטריצת מתארים זמניים V (Y=SV). הרעיון לפירוק וכן משמעות של גורמי הפירוק נלקחו ממודל של temporal decomposition (TD) של אותות דיבור [36] (לפרטים ר׳ פרק 4 בנושא TD בהמשך). הדבר מתבצע ע״י אלגוריתם איטרטיבי, ומתבסס על מזעור מקומי של שגיאה ריבועית משוקללת אדפטיבית, בהינתן אחת המטריצות כתנאי ההתחלה. שיטת פירוק זו קבילה גם עבור קידוד סגמנטים באורך משתנה (ר׳ SegQ בהמשך). המחברים מדווחים על צמצום של כ-30% מהסיביות בהשוואה ל-VQ וצמצום של כ-50% מהסיביות בשילוב עם כימות סגמנטים באורך משתנה. תחום הפעולה של המערכת 300-700 סיביות לשנייה. מערכת הפועלת בקצב קבוע היא בעלת הביצועים של 2.6-1.85 dB (LSD) עבור הקצבים של 300-630 bps בהתאמה.

שיטה נוספת המאפשרת ייעול של MQ הינה Split MQ [22]. כאן מפצלים את מטריצת הפרמטרים הספקטרליים (המכילה וקטורי הפרמטרים הספקטרליים כווקטורי העמודה) למספר תת-מטריצות רוחביות (כלומר בעלות מספר עמודות זהה למטריצת הפרמטרים המלאה). בנוסף, מבצעים חלוקה ל-3 משפחות בהתאם לסוג העירור – כל העמודות שבמטריצה מקורן במסגרות קוליות, אל-קוליות, או מעורבות. כ״א מתת-המטריצות בכל משפחה מקודדות בנפרד ובעזרת מילונים נפרדים. במערכת נוסף גם מנגנון השומר על מונוטוניות הפרמטרים הספקטרליים (להבטחת יציבות במקרה של פרמטרי LSF [9]) וכן מנגנון המונע שינויי ספקטרום חדים מדי, היכולים לגרום לאיכות סובייקטיבית ירודה של הדיבור. הדבר נעשה ע״י אילוץ של מרווח מסוים בין פרמטרי LSF עוקבים בזמן הקידוד. המחברים משתמשים בשגיאה ריבועית משוקללת ייחודית לצורך בניית המערכת וכן במדדים מיוחדים להערכת ביצועי המערכת. כל זאת לאור העדויות, שהמדדים המבוססים על מדידת

עוות ספקטרלי ממוצע (SDM - Spectral Distortion Mean), שהם טובים ומקובלים בהערכות קידוד וקטורי חסר זיכרון, אינם טובים מספיק בקידוד מטריצי [56,2523], כיוון שמדד זה לא מודד חלקות במעבר מבלוק לבלוק. המחברים מדווחים על קידוד באיכות "טובה מאד" בקצבים של כ-650 סיביות לשנייה עם Split MQ.

התקן החדש, הקרוי MELP-1200 [7], אשר מבוסס על התקן MELP-2400 [6], גם הוא מתבסס (חלקית) על עקרון ה-MQ. סכימת הקידוד של וקטורי ה-LSF משלבת כימות וקטורי עם חיזוי לינארי (PVQ) יחד עם קוונטיזציה של שארית החיזוי עבור שתי מסגרות עוקבות במשותף (MQ). ר' סעיף 3.3 לפרטים.

3.6 כימות סגמנטים (SegQ)

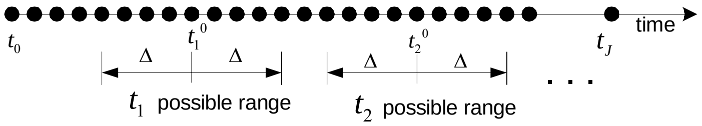
אם מורידים את האילוץ של קצב סיביות קבוע, ניתן אז לדבר על כימות סגמנטים (Segment Quantization, או SegQ), אשר מרחיב את הקידוד המטריצי ומאפשר כימות משותף של מספר משתנה של מסגרות זמניות [32-35]. עדיפות שיטה זאת על MQ נובעת מהמבנה של הדיבור האנושי, אשר מורכב ממקטעי אות קוואזי-סטציונאריים באורכים שונים. מכאן ברור, שהצלחה של SegQ תלויה בשיטת הסגמנטציה של הדיבור לפני (או תוך כדי) פעולת הכימות. בעיה נוספת ב-SegQ הינה הצורך במספר רב יחסית של מילונים לייצוג כל אורך סגמנט אפשרי, דבר שמצריך לימוד ממושך לקבלת ספרי קוד טובים. מקובל גם לאחד את המילונים ע"י הגדרת טרנספורמציה של המרת מימד [330,35], אך הדבר מוסיף עומס חישוב על מערכת הקידוד (בעיית זמן-אמת).

טרנספורמצית המימד ממירה את מימד "הזמן האמיתי" (N) של מטריצת הפרמטרים הספקטרליים (Y, בגודל $p \times N$), המכילה וקטורי פרמטרים ספקטרליים עוקבים מסדר p, למימד "זמן" קבוע M של ספר הקוד. ההמרה נעשית ע"י פעולה לינארית (הכפלה מימין במטריצת טרנספורמציה $N \times M$). הוקטורים המתקבלים הם תוצאה של דגימה במרווחים זהים של המסלול המתקבל מאינטרפולציה לינארית בין וקטורי הפרמטרים הספקטרליים (עמודות של Y) במרחב ה-p-מימדי [33] או, לחילופין, במרחב הזמן בלבד [310,34]. פעולת ההמרה הנ"ל נקראת גם time warping.

Roucos וחבריו [32] הציעו טכניקה לקידוד סגמנטים באורך לא קבוע של ספקטרום הדיבור עבור מקודד דיבור לקצבים נמוכים. הסגמנטציה נעשית בעזרת מדידת שינויים ספקטרליים בין וקטורי הפרמטרים הספקטרליים הסמוכים. ספר הקוד מכיל סגמנטים באורך קבוע והכימות של הסגמנט (באורך משתנה) נעשה בעזרת המרה של סגמנט אותו רוצים לקודד לסגמנט באורך ספר הקוד. שיטה זו אינה ממזערת את השגיאה, כי הסגמנט עובר התמרה לפני ביצוע הכימות. כמו כן ספר הקוד תוכנן ללא קשר לתהליך הסגמנטציה.

ניתן להתגבר על הבעיה הראשונה בעזרת ספר קוד באורך קבוע אשר עובר התמרה לאורך של הסגמנט אותו רוצים לקודד [31,34]. בנוסף ניתן לשלב תהליכי הסגמנטציה וכימות ו/או לתכנן את ספר הקוד במשותף עם אלגוריתם הסגמנטציה [31,34].

ב-[35] Honda ו-Shiraki מתארים את רצף וקטורי ה-LSF בעזרת שרשור של סגמנטים באורך שונה. במערכת זו על המקודד לבחור סגמנטים מייצגים (מתוך ספר קוד, בו כל איבר מייצג רצף של M וקטורי פרמטרים ספקטרליים, ע"י ביצוע time warping), כך שהשגיאה בין רצף וקטורי ה-LSF המקוריים למשוחזרים תהיה מינימאלית, ובנוסף לשדר את אורכי הסגמנטים. מציאת הסגמנטים המייצגים וביצוע הסגמנטציה נעשים בעזרת תכנות דינמי, תוך קביעת סגמנטציה התחלתית, ועידונה. יהי $t_j^0, 0 \le j \le J$ סגמנטציה התחלתית, ויהיו t_j גבולות הסגמנטים, שיש למצוא. המחיר המצטבר של הסגמנט שהסתיים ב- t_j ($t_j^0 - \Delta \le t_j \le t_j^0 + \Delta$) יתקבל בעזרת המחיר של סגמנט שהסתיים ב- t_{j-1} ($t_{j-1}^0 - \Delta \le t_{j-1} \le \min(t_{j-1}^0 + \Delta, t_j)$) ומחיר הכימות של הקטע שבין t_{j-1} ל- t_j. לאחר קביעת סגמנטציה ראשונית מבצעים רקורסיה קדמית למציאת המחיר המצטבר ב- t_j מסוים. ככל ש- Δ גדול יותר תתקבל סגמנטציה טובה יותר (עוות נמוך יותר) במחיר הגדלת הסיבוכיות (ר' איור 3-ב). כמו כן, המחברים מדווחים על מודיפיקציה מוצלחת של האלגוריתם הנ"ל, המאפשרת חפיפות בין הסגמנטים הסמוכים. ברור, כי ההצלחה של השיטה הנ"ל טמונה בין היתר בסגמנטציה התחלתית. שיטה זו (Honda ו-Shiraki) נועדה בעיקרה לאנליזת off-line, והכנסת אילוצי השהיה פוגע בביצועים של השיטה [35].



איור 3-ב מציאת J סגמנטים ע"י תכנות דינמי בהתאם ל-[34]

Figure 3-2. Finding J segments by dynamic programming, according to [34] .

האלגוריתם אשר משלב רעיון של כימות סגמנטים ודילוגים על מסגרות, תוך אילוץ של השהיה וקצב קידוד קבועים פותח באחרונה במעבדה לעיבוד אותות ותמונות בטכניון [30,29]. אלגוריתם זה, המכונה TSQ, מבצע סגמנטציה אופטימלית של בלוק מסגרות קבוע (המכיל 6 מסגרות, כלומר 6 וקטורי פרמטרים ספקטרליים) במובן של שגיאה ריבועית ממוצעת משוקללת מינימלית עם התרת דילוגים בין סגמנטים. וקטורים ספקטרליים שלא משודרים כלל מחושבים במקלט ע"י אינטרפולציה. כמו כן, אורך סגמנט אפשרי יכול לנוע בין 1 ל 4 וקטורי פרמטרים ספקטרליים בלבד ובוחרים בדיוק 4 סגמנטים מתוך כל בלוק. אילוצים אלו מאפשרים לפתור בעיית סגמנטציה וכימות אופטימליים של בלוק מסגרות (בהינתן ספר קוד) בעזרת אלגוריתם תכנות דינמי, אשר ניתן למימוש בעזרת חיפוש

סריג (trellis) בעל 5 דרגות בלבד. אלגוריתם איטרטיבי של סגמנטציה ובניית ספר קוד משפר את הביצועים של השיטה הזאת. הסגמנטים המשודרים קודדו בעזרת Split-MQ באורך 22 סיביות (ז"א קצב שידור ממוצע של 11 סיביות לייצוג פרמטרים ספקטרליים למסגרת). אלגוריתם זה הניב ביצועים דומים לאלו של Split-VQ בקצב של 22 סיביות למסגרת ושימש בהצלחה במקודד 1200 bps מבוסס MELP. יש לציין כי אילוצים של קצב קידוד והשהיה קבועים יחד עם רזולוצית זמן גסה (מסגרות של 22.5 מילישניות) והעדר חפיפות בין בלוקים, מונעים מאלגוריתם זה לאתר מאורעות דיבור אמיתיים (ר' TD בהמשך), ולכן הורדת קצב נוספת תוך שמירה על איכות דיבור סבירה היא בעיתית באלגוריתם זה.

3.7 ייצוג סגמנטים עם מסגרת אנליזה בסיסית באורך משתנה

שיטה נוספת שמתמודדת עם בעיית סגמנטציה והקצאת משאבים מתוארת ב- [36]. ייחודה של השיטה בכך שהיא משלבת בין בעיית ייצוג של מעטפת ספקטרלית לבין בעיית סגמנטציה וכימות. אלגוריתם זה, אשר בוצע ע"י Prandoni ו-Vetterli, מופעל על רצף דגימות זמניות, ומאפשר קביעת אורכים משתנים של מסגרות אנליזה בסיסיות וכן סדר המודל של חיזוי לינארי (LP) בצורה אופטימלית, במובן של קצב-עוות. (כל השיטות האחרות שהוזכרו לעיל פועלות עם מסגרת אנליזה בסיסית באורך קבוע). המטרה של השיטה – למצוא סגמנטציה **t** (אוסף קצוות סגמנטים, כלומר $\mathbf{t}=\left\{t_0\triangleq 0<t_1<t_2<\cdots<t_{k-1}<t_k\triangleq N\right\}$) וייצוג של פרמטרים ספקטרליים **p** (אוסף סדרי חיזוי לינארי המתאימים לכל סגמנט) עבור בלוק באורך N $\left(x=\left\{x(0)\ldots x(N-1)\right\}\right)$ בצורה אופטימלית, במובן של מינימום MSE, עם אילוץ של קצב שידור. לשם כך מבצעים מזעור של ה-MSE על סט סגמנטציות $T_{[0,N]}$ וכן קבוצת מודלי חיזוי לינארי בסדרים שונים P, תוך האילוץ הנ"ל, כלומר

$$\begin{cases}\min\limits_{t\in T}\min\limits_{p\in P(t)}\left\{D(x,t,p)\right\}\\ R(x,t,p)\le R_c\end{cases}\qquad(3.2)$$

כאשר,

$$D(x,t,p)=\sum_k d_x^2(t_k,t_{k+1};p_k)$$
$$R(x,t,p)=\sum_k r(p_k)$$

וכן $d_x^2(t_k,t_{k+1};p_k)$ הוא ה-MSE של סגמנט (t_k,t_{k+1}), בייצוג של מודל LP מסדר p_k, ו- $r(p_k)$ מציין את מספר הסיביות המשמשות לקידוד מודל מסדר p_k, כולל מידע צד על אורך סגמנט.

הבעיה ניתנת לפתרון ע"י כופלי לגראנז' בצורה הבאה:

1) מציאת לגרנז'יאן עבור λ_0 שרירותי -

$$\begin{cases} J^*(\lambda_0) = \min_{t\in T} \min_{p\in P(t)} \{J(\lambda_0)\} \\ J(\lambda) = D(x,t,p) + \lambda R(x,t,p) \end{cases} \quad (3.3)$$

2) עדכון של $\lambda = \lambda_i$, תוך ניצול העובדה ש- $J(\lambda)$ הינה פונקציה מונוטונית לא עולה (הגדלה של λ פירושה הקטנה של קצב שידור). לדוגמה, ע"י שיטת החצייה.

3) חזרה על חישוב הלגרנז'יאן לפי (3.3) ועדכון λ עד לקבלת קצב שידור רצוי.

חישוב של (3.3) ניתן לביצוע יעיל ע"י תכנות דינמי:

$$J^*_{[0,t]}(\lambda) = \min_{0\le\tau\le t-1} \left\{ J^*_{[0,\tau]}(\lambda) + \min_{1\le p\le Q} \left\{ d_x^2(\tau,t;p) + \lambda r(p) \right\} \right\} \quad (3.4)$$

נסכם את שלבי האלגוריתם:

i. יצירת סריג, כאשר כל מצב $s_{t,v}$ בו מייצג מסגרת אנליזה בזמנים $[c(t-v), c(t+1)-1]$, וקיימת קשת בין כל $s_{t,v}$ ל- $s_{t+1,v+1}$. (הרזולוציה בציר הזמן של האלגוריתם הנה c דגימות, ז"א קביעת קצוות הסגמנטים נעשית בקפיצות של c דגימות). ר' איור 3-ג להדגמה.

ii. בכל מצב נשמרים ערכי מקדמי האוטוקורלציה וה-MSE (לכל אחד מסדרי אנליזת LP המותרים).

iii. לכל שלב S_t מחשבים מחיר לגראנז' מצטבר מינימלי

$$J^*_{t+1} = \min_v \min_i \left\{ J^*_{t-v} + d_y^2(t-v, t+1; p_i) + \lambda r(p_i) \right\} \quad (3.5)$$

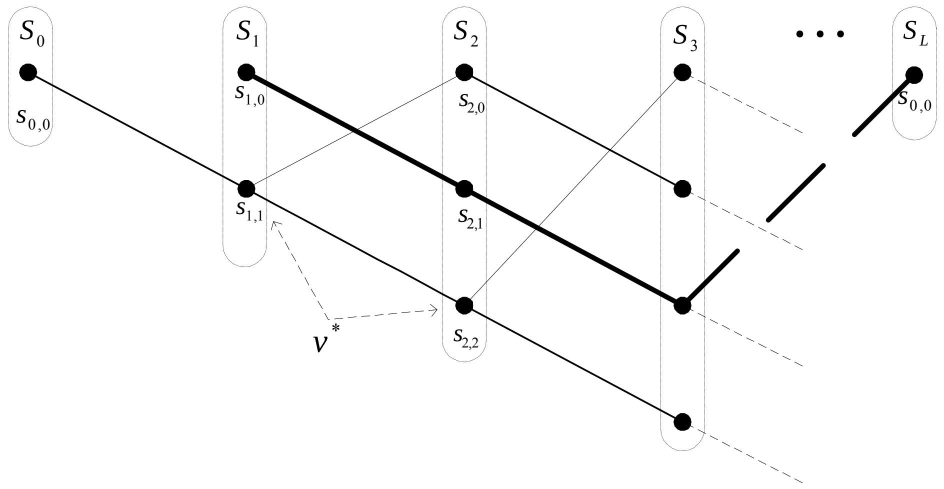
ומשייכים אותו לקשת בין s_{t,v^*} ל- $s_{t+1,0}$, כאשר:

$$v^* = \arg\min_v \min_i \left\{ J^*_{t-v} + d_y^2(t-v, t+1; p_i) + \lambda r(p_i) \right\} \quad (3.6)$$

iv. אוספים רקורסיבית i^*, v^* לאורך המסלול שהסתיים ב- S_L, וכך מוצאים את הסגמנטציה ואוסף סדרי אנליזת LP אופטימליים (ר' איור 3-ג).

v. ע"מ לקבוע נקודת עבודה אופטימלית, יש לחזור על תהליך ולקבוע λ המתאים לקצב הנדרש. ניתן לעשות זאת בשיטת החצייה כיוון ש $J(\lambda)$ הינה פונקציה מונוטונית.

בשיטה זו הורדת קצב השידור גורמת לתופעה של ייצוג מקטעים ארוכים יחסית ע"י וקטור פרמטרים ספקטרליים יחיד, דבר שפוגע בדינמיקה של הדיבור. ניתן להקטין את חומרת התופעה ע"י שיטות אינטרפולציה בתוך מסגרות ארוכות, אך קשה למגר את התופעה.



איור 3-ג. . דיאגרמת הסריג עבור האלגוריתם הדינמי. הסגמנטציה האופטימלית כאן היא {0, 1, L}. (הקצוות של הסגמנטים הם שייכים לאוסף $\{i \mid s_{i,0}$ is on the route, ending in $S_L\}$.

Figure 3-3. Trellis diagram of the dynamic programming algorithm. The optimal segmentation here is {0, 1, L}. (The edges of segments are in the set $\{i \mid s_{i,0}$ is on the route, ending in $S_L\}$).

3.8 סיכום

בפרק זו הצגנו מגוון רחב של שיטות להורדת היתירות הזמנית של הפרמטרים הספקטרליים של אות הדיבור. רוב השיטות הנ"ל השתמשו באחת משתי הגישות הבאות: הדילוג על מסגרות או ייצוג משותף של מספר מסגרות. תכן של המקודדים בגישות הללו נעשה קשה עד בלתי אפשרי לקצבים של כ-300 bps, וזאת מכיוון שדילוג על מסגרות בקצבים אלה מביא לדגרדציה משמעותית באיכות ועבור ייצוג משותף של מספר הוקטורים (שיטות, כגון MQ, SegQ) קיים קושי רב בתכן ספרי קוד כלליים (speaker independent) בקצבים האלה ללא דגרדציה משמעותית בביצועים. בפרק הבא נציג גישה שונה לייצוג וקוונטיזציה של הפרמטרים הספקטרליים, המתבססת על פירוק אות הדיבור לרצף של מאורעות זמניים ונקראת Temporal Decomposition.

פרק 4

מודל הפירוק למאורעות זמניים

(Temporal Decomposition)

4.1 מבוא

שיטת ייצוג של אות הדיבור ע״י פירוק למאורעות זמניים, הקרויה Temporal Decomposition (TD), הוצעה לראשונה ע״י Atal [37]. מאז הפרסום הראשון ב-1983 נעשו הרבה עבודות המשתמשות במודל זה לצורכי ייצוג, סגמנטציה וקוונטיזציה של דיבור. [37-520]. בשיטה זו חותרים לייצוג של מעטפת ספקטרלית בדידה (המיוצגת ע״י מטריצת הפרמטרים הספקטרליים **Y** במימדים $p \times N$ (3.1) של מקטע דיבור) ע״י אוסף של M $(M < N)$ *מאורעות*. כל מאורע מיוצג ע״י וקטור פרמטרים ספקטרליים אחד (הקרוי וקטור המטרה) ופונקצית אינטרפולציה זמנית (המכונה *פונקצית מאורע*). פונקציות מאורע זו חייבת להיות קומפקטית (שונה מאפס רק בקרבת וקטור המטרה) אך פונקציות המאורעות יכולות לחפוף ביניהן. כל מאורע מאופיין ע״י וקטור מטרה, פונקצית המאורע הקומפקטית ו*מרכז המאורע* – שהוא הזמן איתו מזוהה בד״כ וקטור המטרה והוא נמצא באזור מרכזי של פונקצית המאורע. האוסף של מרכזי המטרה ייקרא בעבודה הנוכחית *הסגמנטציה*, כיון שהדיבור מפורק לסגמנטים באורכים שונים. יש להעיר, שמכיוון שמודל ה-TD מופעל על וקטורי הפרמטרים הספקטרליים הדוגמים את המעטפת הספקטרלית הרציפה של אות הדיבור, הזמן נמדד גם הוא ביחידות של מסגרת האנליזה, כלומר הזמן במודל ה-TD אקוויוולנטי למספר הסידורי של העמודה המתאימה במטריצת הפרמטרים הספקטרליים **Y**. מודל ה-TD מתקשר למבנה של דיבור אנושי המורכב מרצף של תנועות ארטיקולטוריות (articulatory movements) המעבירות את כלי הקול ממצב אחד לשני, וכך מייצרות את מאורעות הדיבור (פונמות) [37].

4.2 המודל הכללי של TD

4.2.1 תיאור כללי

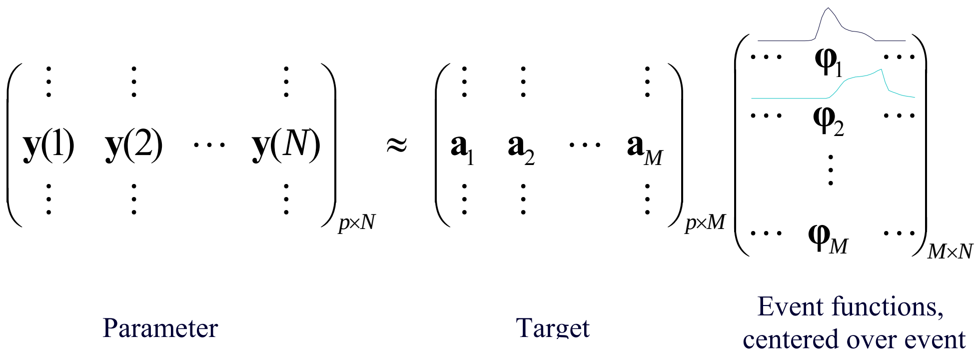
בייצוג זה כל וקטור פרמטרים ספקטרליים בקטע באורך N מסגרות מבוטא ע״י:

$$\mathbf{y}(n)=\sum_{k=1}^{M}\mathbf{a}_k\phi_k(n),\ 1\le n\le N\ , \tag{4.1}$$

כאשר $\mathbf{a}_k$ מציין וקטור מטרה ו- $\phi_k(n)$ את ערך פונקצית המאורע ה-k-ית במסגרת n. ברישום מטריצי:

$$\mathbf{Y}=\mathbf{A\Phi}\ , \tag{4.2}$$

כאשר, מטריצת פונקציות המאורע $\mathbf{\Phi}$ הינה מטריצה דלילה (כיוון שפונקציות המאורע שונות מאפס רק בקרבת מיקום המאורע) במימד $M\times N$ המכילה פונקציות אינטרפולציה כוקטורי שורה, ומטריצת וקטורי המטרה (יעדים) $\mathbf{A}$ הינה מטריצה $p\times M$ המכילה וקטורי מטרה כוקטורי עמודה, כאשר כל וקטור מטרה כזה מכיל p פרמטרים ספקטרליים המייצגים את מעטפת הדיבור במסגרת בודדת (ר' איור 4-א). איור סכימתי של פעולת ה-TD ניתן לראות באיור 4-ג.

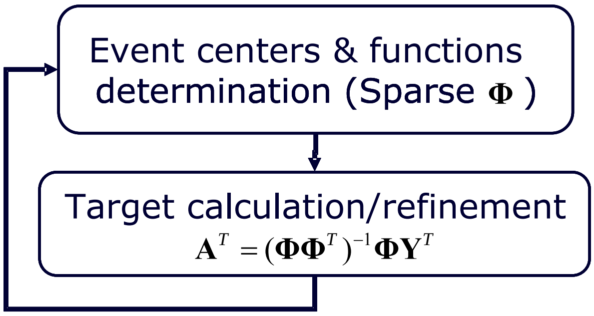


איור 4-א. סימונים עבור מודל הפירוק למאורעות זמניים (Temporal Decomposition).

Figure 4-1. Temporal Decomposition notation.

כיוון שאין אילוצי דלילות על מטריצת וקטורי המטרה $\mathbf{A}$, ניתן למצוא את ערכה האופטימלי במובן של סכום ריבועים מינימלי (Least Squares), בהנחה שידועה מטריצת ה-$\mathbf{\Phi}$, כפי שיוצג בהמשך הפרק. לעומת זאת, אין דרך מוגדרת אחת למציאת מטריצת $\mathbf{\Phi}$ כך שתהיה דלילה (כלומר פונקציות המאורע, שהן שורותיה, יהיו בעלי תמך מצומצם). למעשה, רוב השיטות של פירוק ה-TD נבדלות בעיקר בדרכים ובשיטות לחישוב מטריצת ה-$\mathbf{\Phi}$ [43,40,39,370]. תהליך זה צריך במקביל למצוא את מרכזי המאורעות (או להניח הנחות לגביהם).

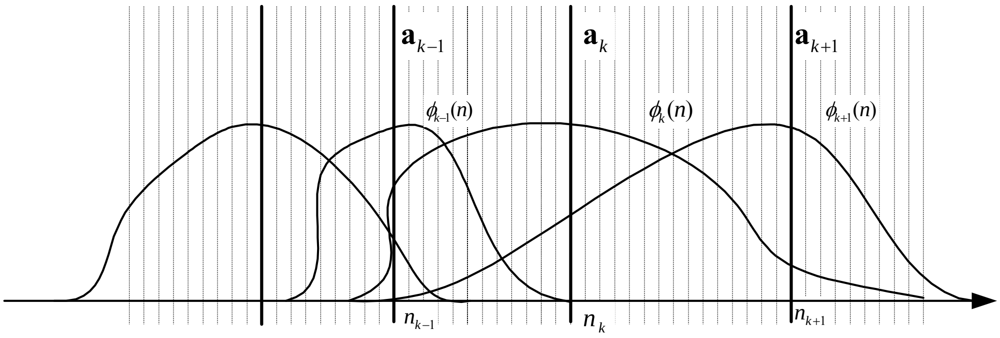
תהליך ה-TD הכללי מורכב בד"כ משני שלבים עיקריים: מציאת פונקציות המאורעות וחישוב וקטורי המטרה. ניתן לבצע יותר מאיטרציה אחת של האלגוריתם לשיפור איכות ההתאמה (ר' איור 4-ב).



איור 4-ב. שלבי ה-TD הכללי.

Figure 4-2. General TD stages.

סוגיה חשובה ב-TD היא בחירת הפרמטרים הספקטרליים עליהם תתבצע הסגמנטציה. האלגוריתם המקורי [37] משתמש ב-Log-Area coefficients (LA) בתור פרמטרים ספקטרליים. ב-[46] נטען כי Log-Area ratios (LAR) מניבים את התוצאות הטובות ביותר בעת שחזור הספקטרום. במחקר נוסף לבחינת פרמטרי הספקטרום לשם TD נטען כי LA טובים לזיהוי המאורעות, ו-LAR לעידון מיקומם ושחזור הספקטרום [50]. מקדמי LSF, כאמור, ידועים בתכונות אינטרפולציה וקוונטיזציה טובות שלהם [10,9], מה שמקל על קידוד של וקטורי המאורע (הם בד"כ דומים מאד לוקטורי הפרמטרים הספקטרליים שיש לקודדם). באחרונה, מספר רב של אלגוריתמי TD נוסו והותאמו למקדמי ה-LSF [4341,39].



איור 4-ג. פירוק ה-TD הכללי.

Figure 4-3. General Temporal Decomposition model.

4.2.2 מציאת מטריצת וקטורי המטרה

בסעיף זה נתאר דרך יעילה לחישוב מטריצת וקטורי המטרה בהינתן מטריצת פונקציות המאורעות $\mathbf{\Phi}$. נחפש פתרון אופטימלי במובן של סכום ריבועים מינימלי (LS), כלומר

$$\mathbf{A}_{opt} = \arg\min_{\mathbf{A}} \left(\mathrm{tr}\left((\mathbf{Y} - \mathbf{A\Phi})^T (\mathbf{Y} - \mathbf{A\Phi}) \right) \right) \tag{4.3}$$

גזירה של הביטוי לפי $\mathbf{A}$ והשוואה לאפס מביאה למשואה מטריצית הבאה:

$$(\mathbf{\Phi\Phi}^T)\mathbf{A}^T = \mathbf{\Phi Y}^T \tag{4.4}$$

יהיו $[\mathbf{A}^T]_i, [\mathbf{Y}^T]_i$ העמודות ה-i במטריצות $\mathbf{A}^T$ ו- $\mathbf{Y}^T$, בהתאמה. נפרק את משואה (4.4) ל-p מערכות משואות לינאריות, המתאורות מסלול ההשתנות של הרכיב ה-i-י לאורך זמן:

$$(\mathbf{\Phi\Phi}^T)[\mathbf{A}^T]_i = \mathbf{\Phi}[\mathbf{Y}^T]_i, \quad 1 \le i \le p \tag{4.5}$$

כיוון שהמטריצה $\mathbf{\Phi\Phi}^T$ הנה סימטרית ודלילה (מרביתה אפסים, פרט למספר קטן של אלכסונים מרכזיים השונים מאפס, בהתאם לתכונות החפיפה של פונקציות המאורעות), ניתן לפתור כ"א מ-p מערכות המשוואות הנ"ל בדרכים יעילות [58]. בפרט, אם מותרת חפיפה של פונקציות המאורע השכנות בלבד, המטריצה $\mathbf{\Phi\Phi}^T$ היא תלת אלכסונית (tri-diagonal). (נראה בסעיפים הבאים שזהו המקרה שמעניין אותנו במיוחד, כיוון שהוא מגדיר את מודל פירוק ה-TD המצומצם, שנתבסס עליו בהמשך.) במקרה זה נוכל לפתור כ"א ממערכות המשוואות (6.4) ע"י Simultaneous symmetric Gaussian elimination [580]. בהמשך הפרק נתאר מימושים ספציפיים לפעולת ה-TD, אשר נבדלים, בעיקר, ביישום השלב של מציאת מטריצת פונקציות המאורע.

4.3 שיטות פירוק המבוססות על SVD

השיטה הקלאסית המוצעת ע"י Atal [36] וכן שיטות נוספות [46] ל-TD, מתבססות בעיקרן על מציאת וקטורים וערכים עצמיים (ע"ע) באמצעות פירוק SVD [58], דבר שהופך את השיטות הללו ליקרות במיוחד מבחינה חישובית (סיבוכיות של השיטה היא $O(n^4)$). השיטה המקורית של Atal מכילה את השלבים הבאים:

שלב I. מציאת פונקציות מאורע ראשוניות

יהי $\mathbf{Y}$ מטריצת $p\times T$ של פרמטרים ספקטרליים (Atal מציע להשתמש בפרמטרי LA [1]) של משפט שלם (או מקטע ארוך שמשני צידיו קטעי שקט). מעוניינים למצוא פירוק TD ל-$\mathbf{Y}$ (4.2). לשם כך מבצעים כל פעם אנליזה של תת-מטריצה שלה $\mathbf{Y}_t$ בגודל $p\times N$, $N\ll T$:

$$\mathbf{Y}_t=\left[\mathbf{y}_{t-N/2}\quad\cdots\quad\mathbf{y}_t\quad\cdots\quad\mathbf{y}_{t+N/2}\right], \tag{4.7}$$

כאשר $\mathbf{y}_t$ הוא וקטור עמודה ה-t-י של $\mathbf{Y}$.

בוחרים N גדול מספיק, כך שדרגת מטריצת $\mathbf{Y}_t$ לא תהיה קטנה ממספר המאורעות M שעשויים להימצא בתוכה (Atal מציע לעבוד על מקטעים באורך של 200-300 ms ולבחור $M=5$). במקרה כזה אם קיימת פונקצית מאורע הממורכזת סביב t (שורה ממטריצה $\mathbf{\Phi}$, שאזור התמך שלה ממורכז סביב t ואינו גדול מ-N) אז היא צירוף לינארי של השורות במטריצה $\mathbf{Y}_t$ או, לחילופין, צירוף לינארי של M וקטורים עצמיים המתאימים ל-M ערכים עצמיים (ע"ע) הגדולים ביותר של המטריצה $\mathbf{Y}_t$.

לכן, לכל $n\in\mathbb{Z}$, $t=\Delta\cdot n$, $N/2\le t\le T-N/2$, כאשר Δ היא גודל הצעד (ברזולוציה של מסגרות):

1) מבצעים פירוק SVD [58] למטריצת הפרמטרים הספקטרליים $\mathbf{Y}_t$:

$$\mathbf{Y}_t^T=\mathbf{UDV}^T, \tag{4.8}$$

כאשר $\mathbf{D}$ היא אלכסונית ו-$\mathbf{U}$ ו-$\mathbf{V}$ הן אורתוגונליות. $\mathbf{U}$ מכילה וקטורים עצמיים של $\left(\mathbf{Y}_t^T\mathbf{Y}_t\right)$ כעמודותיה ו- $\mathbf{V}^T$ מכילה וקטורים עצמיים של $\left(\mathbf{Y}_t\mathbf{Y}_t^T\right)$ כשורותיה. בשלב ראשון בוחרים M וקטורים עצמיים (עמודות של $\mathbf{U}$) המתאימים ל-M ע"ע הגדולים ביותר, שבעזרתם שואפים לבטא את פונקצית המאורע הקרובה לרגע t:

$$\tilde{\boldsymbol{\varphi}}_t=\mathbf{b}_t^T\begin{bmatrix}\mathbf{u}_1^T\\ \vdots\\ \mathbf{u}_M^T\end{bmatrix},\quad \mathbf{b}_t\triangleq\left[b_{t,1}\quad\cdots\quad b_{t,M}\right], \tag{4.9}$$

כאשר $\left\{\mathbf{u}_i\right\}_{i=1}^M$ הינם וקטורים עצמיים המתאימים ל-M הערכים העצמיים הגדולים ביותר של המטריצה $\mathbf{Y}_t$, $\left\{b_{t,i}\right\}_{i=1}^M$ הנם סקלרים שיש לחשב אותם ו- $\tilde{\boldsymbol{\varphi}}_t$ היא פונקצית המאורע הממורכזת סביב t. או, ברישום סקלרי:

$$\phi_t(t-\tfrac{N}{2}+n)=\sum_{i=1}^{M} b_{t,i}u_i(n), \quad 0\le n\le N-1 \tag{4.10}$$

לאחר מכן, פונקצית המאורע $\phi_t(n)$ (הקרובה ביותר לנקודה מרכזית t) מחושבת ע״י מינימיזציה של מידת מרחק, $\theta(t)$:

$$\theta(t)=\sqrt{\frac{\sum_{n=-N/2}^{N/2-1}(n-t)^2\phi_t^2(n)}{\sum_{n=-N/2}^{N/2-1}\phi_t^2(n)}} \tag{4.11}$$

מידת מרחק זו מהווה מדד לקומפקטיות של פונקצית המאורע (סטיית התקן של משתנה אקראי בעל צפיפות ההסברות של $\phi_t^2(n)$ המנורמלת). ניתן להראות [36], כי המינימיזציה של (4.11) ביחס ל-**b** שקולה לבעיית הערכים העצמיים (ע״ע) :

$$\mathbf{Rb}=\lambda\mathbf{b} \tag{4.12}$$

כאשר $1\le i,r\le M$, $R_{ir}=\sum_{n=1}^{N}(n-n_c)^2u_i(n)u_r(n)$. פתרון, המתאים לע״ע הקטן ביותר, מביא ל-**b** אופטימלי.

2) מיקום מדוייק של מרכז של מאורע $\phi_t(n)$ (c_t) ייקבע כנקודה בה $v(l)$ חוצה את קו האפס מלמעלה למטה, כאשר $v(l)$ מוגדר ע״י :

$$v(l)=\frac{\sum_{n=t-N/2}^{t+N/2-1}(n-l)\phi_t^2(n)}{\sum_{n=t-N/2}^{t+N/2-1}\phi_t^2(n)} \tag{4.13}$$

<u>שלב II. עידון פונקציות המאורע</u>

בשלב זה חוזרים על החישוב של <u>שלב I</u> רק עבור הנקודות c_t, כאשר משתמשים באורך תת-מטריצה $\mathbf{Y}_{c_t}$ משתנה, כך שתכיל בדיוק M מאורעות (למעט מאורעות קצה). כך, למעשה מבצעים עידון והפחתה של כמות המאורעות. שלב זה יכול להכיל כמה איטרציות זהות.

<u>שלב III. חישוב וקטורי המאורע</u>

לאחר שפונקציות המאורע נקבעו, מוצאים את **A** ע״י מינימיזציה של שגיאה ריבועית (SE) :

$$E_i = \sum_{n=1}^{N}\left(y_i(n) - \sum_{k=1}^{M} a_{ik}\phi_k(n) \right)^2, \quad 1 \le i \le P \tag{4.14}$$

בעיה זו שקולה לבעיית ה-LS , המתוארת בסעיף 4.2.2.

<u>שלב IV. שיפור צורת פונקציות המאורע</u>

לבסוף, ניתן לעדן את התוצאה ע"י אלגוריתם איטרטיבי של מציאת $\mathbf{\Phi}$ ו $\mathbf{A}$ לסירוגין, באמצעות מינימיזציה של (4.14), כאשר בכל איטרציה מחשבים פונקציות מאורע אופטימליות (בהתאם ל-(4.14)), מאפסים אונות הצד של שלהם (כך שתישאר אונה ראשית בלבד) ומחשבים מקדמים אופטימליים עבור פונקציות המאורע המתוקנות.

כפי שניתן להיווכח, אלגוריתם זה אינו ישים במערכות זמן אמת עקב העומס החישובי הרב שהוא מצריך וההשהיה הגבוהה הכרוכה בו. כמו כן, ניכרת בו תלות חזקה בגודל הצעד Δ, אורך תת-המטריצה $\mathbf{Y}_t$, מספר הע"ע הנבחרים M ועוד. ע"מ להפוך אותו לישים ושימושי לקידוד דיבור, יש צורך להימנע מפעולת ה-SVD (שחוזרים עליה לעתים תכופות באלגוריתם המקורי).

4.4 שיטות מודרניות לביצוע TD בהנחת סגמנטציה התחלתית

באחרונה הוצעו מספר שיטות לפירוק TD שהן ישימות מבחינה חישובית ומתאימות יותר לצרכי קידוד [39-44]. יתרון הנוסף של השיטות הללו, לשימושי קידוד דיבור, שהן נוסו בהצלחה על פרמטרי ה-LSF, שתכונותיהם הנוחות לקידוד הוזכרו בתחילת העבודה. השיטות הללו ממקמות את המאורעות בנקודות היציבות המקומית של ספקטרום הדיבור, עפ"י עקרון יציבות ספקטרלית [39,39], במקום שימוש ב-SVD, או לחילופין מבצעים חיפוש של מיקומי המאורעות המיטביים על פני קטע דיבור קצר [44].

בעבודות [43,041,40,39], וקטורי המטרה משוערכים תחילה כוקטורי פרמטרים ספקטרליים (LSF) הממוקמים בנקודות יציבות ספקטרלית. נקודות אלו נקבעות ע"ס נקודות מינימום מקומי של פונקצית קצב ההשתנות של פרמטרים ספקטרליים (Spectral feature Transition Rate (SFTR)). פונקציה זו מוגדרת בנקודה n באמצעות שקלול שיפועים של קווי רגרסיה של פרמטרים ספקטרליים בסביבה קרובה של הנקודה:

$$\begin{aligned} c_i(n) &= \frac{\sum_{m=-R}^{R} m y_i(n+m)}{\sum_{m=-R}^{R} m^2}, \quad 1 \le i \le p \\ SFTR:\ s(n) &= \sum_{i=1}^{P} c_i(n)^2, \qquad 1 \le n \le N \end{aligned} \tag{4.15}$$

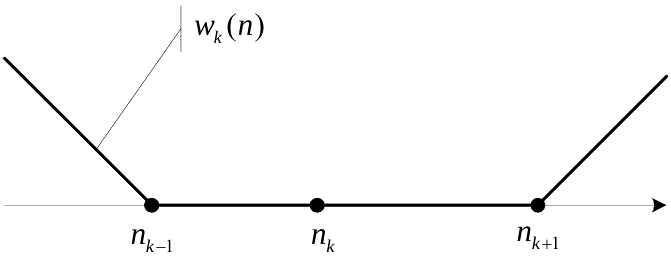
כאשר R נבחר להיות שווה ל-2. ראוי לציין כי קריטריון זה מתאים לוקטורי ה-LSF בגלל תכונות ההשתנות האיטית שלהם.

הרעיון לקבוע את וקטורי המאורע באזורי היציבות (כלומר הוקטורים נמצאים בסביבות סטציונריות) יוצא מנקודת ההנחה כי ניתן יהיה לתאר את אזורי המעבר בין הסגמנטים הסטציונריים ע"י פונקציות המאורע (החופפות). שיטה זו עלולה להיכשל באיתור מאורעות קצרים (כמו עיצורים בתוך רצף קולי) וכן מאורעות אשר באים לידי ביטוי בהשתנות זמנית ולא השתנות ספקטרלית (דוגמת עיצורים פוצצים). ואכן כך הדבר עבור שיטת RTD [39,40], שמתבססת גם היא על קביעת מיקומי המאורעות ע"י SFTR. (האלגוריתם יפורט בהמשך סעיף זה.) נמצא, כי SFTR המופעל על LSF אינו מגלה יותר מאשר כ-10 מאורעות לשנייה ולכן מושתלים מאורעות נוספים בנקודות בהם יש לשגיאת השחזור מקסימום מקומי העולה על ערך סף נתון [41].

בשני הסעיפים הבאים נתאר שתי שיטות פירוק TD, אשר מתבססות על הבחירה ההתחלתית של מרכזי המאורעות באמצעות ה-SFTR. מרכזי המאורעות קובעים גם ערך התחלתי של וקטורי המטרה, שכן אלה נבחרים להיות וקטורי הפרמטרים הספקטרליים עצמם במרכזי המאורעות. הייחוד של השיטות הללו טמון בדרך מציאת פונקציות המאורעות בהינתן וקטורי המטרה ומיקומי מרכזי המאורעות.

4.4.1 הגבלת התמך של פונקציות במאורעות ע"י כופלי לגראנז'

Nandasena *et al* [39] מציעים שיטה איטרטיבית לקביעת פונקציות המאורע, בהינתן וקטורי המטרה ההתחלתיים. טכניקה זו מאלצת תמך מצומצם של פונקציות מאורע. פונקציות המאורע נקבעות ע"י מינימיזציה של שגיאה ריבועית יחד עם משקל פונקצית המאורע, אשר בא לידי ביטוי מחוץ לסביבה הקרובה התחומה ע"י *מרכזי המאורעות* השכנים. המשקל הולך ועולה לינארית ככל שמתרחקים ממרכזי המאורעות השכנים, וכך מגביל אפקטיבית את משך פונקצית המאורע (ר' איור 4-ד).



איור 4-ד. פונקצית משקל עבור המאורע ה-*k*-י.

Figure 4-4. Weighting function for the *k*-th event

הלגרנז'יאן שיש למזער במקרה זה:

(4.16)

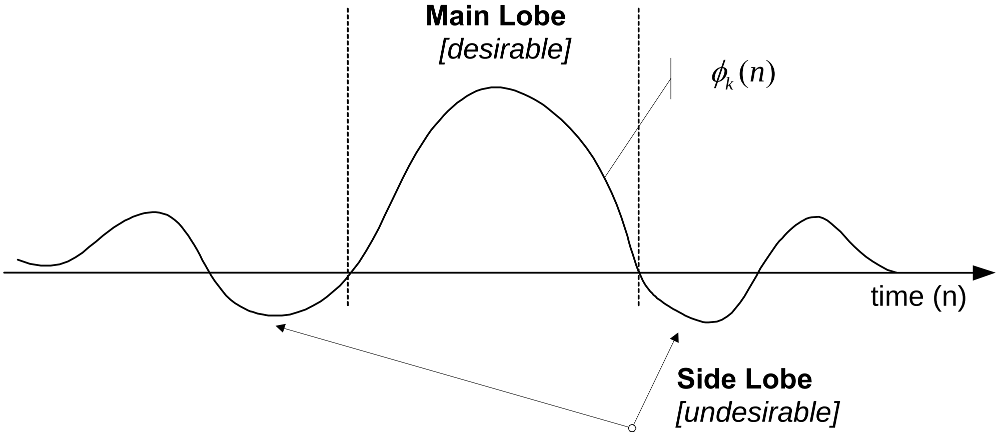
$$J(\boldsymbol{\varphi}(n),\lambda)=\sum_{i=1}^{P}(y_i(n)-\hat{y}_i(n))^2+\lambda\sum_{k=1}^{M}w_k(n)^2\phi_k(n)^2,\ \boldsymbol{\varphi}(n)=\begin{bmatrix}\phi_1(n) & \phi_2(n) & \cdots & \phi_M(n)\end{bmatrix}^T$$

והפתרון ניתן בצורה מטריצית:

$$(4.17)\quad \boldsymbol{\varphi}(n)=(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{W}_{\mathbf{n}}^{\mathrm{T}}\mathbf{W}_{\mathbf{n}})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}(n)\ ;\qquad \mathbf{W}_{\mathbf{n}}=\begin{pmatrix} w_1(n) & 0 & \mathbf{0} \\ 0 & \ddots & 0 \\ \mathbf{0} & 0 & w_M(n)\end{pmatrix}$$

לאחר קביעות אילו, מקבלים פונקציות מאורעות בעלות אונה ראשית ומספר אונות צד (ר' איור 4-ה). כעת מריצים אלגוריתם איטרטיבי להורדת אונות הצד שבמהלכו חוזרים על 2 צעדי עדכון: ראשית חוזרים על פעולת המינימיזציה בהתאם ל-(4.16) ו-(4.17), עם פונקציות משקל $w_k(n)$ מתוקנות, שגדלות מעבר לאונות צד של פונקציות המאורע $\phi_k(n)$ (איור 4-ה). לאחר מכן מוצאים את וקטורי המטרה ע"י מינימיזציה של (4.14) בהתאם לסעיף 4.2.2. האלגוריתם עוצר, כאשר האנרגיה באונות צד היא מתחת לסף נתון יחסית לאנרגית האונה הראשית.

אלגוריתם זה הינו יעיל יותר ויציב יותר מהאלגוריתם המקורי של Atal. האלגוריתם מאפשר חפיפות של מספר לא מוגבל של מאורעות. ברור, כי המזעור של (4.16) אינו מביא בהכרח למציאת פונקציות מאורע ווקטורי המטרה האופטימליים, במובן של מינימום שגיאת שחזור ריבועית, עקב הצורך באילוץ, כמו כן ניכרת התלות בקביעה התחלתית של מיקום מרכזי המאורעות.

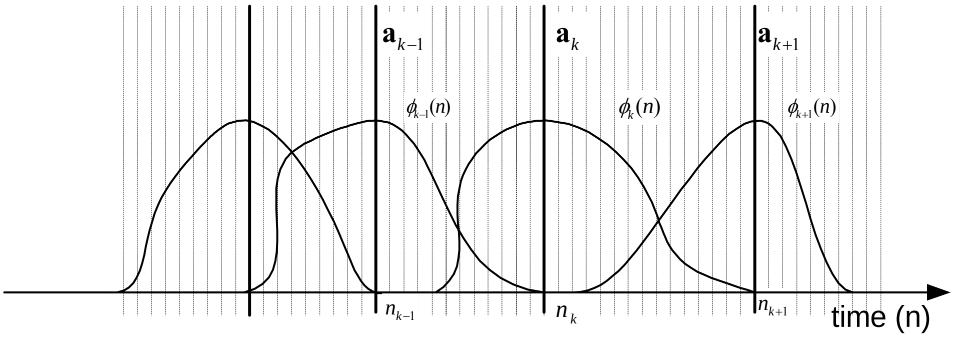


איור 4-ה. . הצורה האופיינית של פונקצית מאורע התחלתית. ניתן לראות אונות צד משניות, אשר אינן רצויות.

Figure 4-5. Typical shape of an initial event function. Note the presence of undesirable side lobes.

4.4.2 אילוץ קשיח של תמך המאורעות (RTD)

שיטה נוספת לחישוב פונקציות מאורע, המכונה Restricted TD ‏(RTD) הוצעה ע"י Kim *et al* [40,41]. (שיטה דומה הוצעה במקביל ע"י Athaudage *et al* [43] ומכונה שם TD מסדר שני). בשיטה זו קיים אילוץ של ממש על חפיפת פונקציות מאורעות שכנים בלבד (ר' איור 4-ו).



איור 4-ו. פירוק ה-TD המצומצם (RTD).

Figure 4-6. Restricted Temporal Decomposition (RTD) model of speech.

המודל של TD עם אילוץ של חפיפת מאורעות שכנים בלבד נתונה ע"י:

$$\hat{\mathbf{y}}(n) = \mathbf{a}_k \phi_k(n) + \mathbf{a}_{k+1}\phi_{k+1}(n), \quad n_k \le n < n_{k+1} \tag{4.18}$$

כאשר $\hat{\mathbf{y}}(n)$ הוא וקטור הפרמטרים הספקטרליים המקורב במסגרת n, ו-n_k, n_{k+1} הם מיקומי המאורעות $k, k+1$ בהתאם (ר' איור 4-ו). בהנחה שידועים מיקומי המאורעות הראשוניים, ניתן למזער שגיאת הקירוב בכל רגע n, המוגדרת ע"י:

$$E(n) = \left\| \mathbf{y}(n) - \hat{\mathbf{y}}(n) \right\|^2 \tag{4.19}$$

הצבה של (4.18) לתוך (4.19) ומינימיזציה ע"י גזירה לפי והשוואה לאפס מביאה לפתרון אופטימלי לפונקציות המאורע רגעיות (בהינתן מרכזי המאורעות ובהנחות ה-RTD, כמובן) [44]:

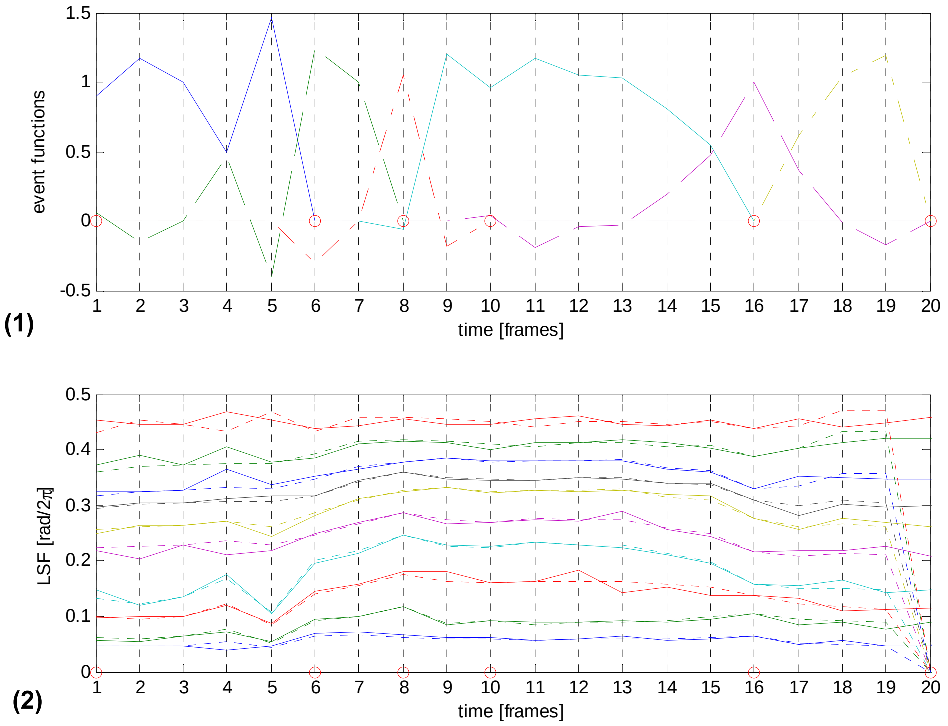
$$\begin{pmatrix} \phi_k(n) \\ \phi_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_k^T \mathbf{a}_k & \mathbf{a}_k^T \mathbf{a}_{k+1} \\ \mathbf{a}_k^T \mathbf{a}_{k+1} & \mathbf{a}_{k+1}^T \mathbf{a}_{k+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_k^T \mathbf{y}(n) \\ \mathbf{a}_{k+1}^T \mathbf{y}(n) \end{pmatrix}, \quad n_k \le n < n_{k+1} \tag{4.20}$$

בהעדר אינפורמציה אחרת, פרט למרכזי המאורעות, מקובל לקבוע, כי וקטור המטרה שווה לווקטור הפרמטרים שבמרכז המאורע, כלומר:

$$\mathbf{a}_k = \mathbf{y}(n_k) \tag{4.21}$$

הנחה זו יחד עם (4.20) מביאה לפתרון אנליטי סגור לפונקציות מאורע אופטימליות.

ניתן לראות (איור 4-ז), כי צורת פונקציות מאורע אופטימליות עלולה להיות אי-רגולריות, דבר שיכול להקשות על הקוונטיזציה שלהן.



איור 4-ז. פירוק RTD עם פונקציות מאורע אופטימליות (לא מאולצות). (1). פונקציות מאורעות לא מאולצות עבור פירוק RTD עם סגמנטציה נתונה (כלומר מרכזי מאורעות ידועים). מרכזי המאורעות מסומנים בעיגולים. (1). קירוב ל-LSF, אשר התקבל ע"י פירוק ה-RTD המאויר ב-(1). ה-LSF המקוריים מסומנים בקווים רציפים, וה-LSF המודללים מסומנים בקווים מקווקווים. מרכזי המאורעות מסומנים בעיגולים.

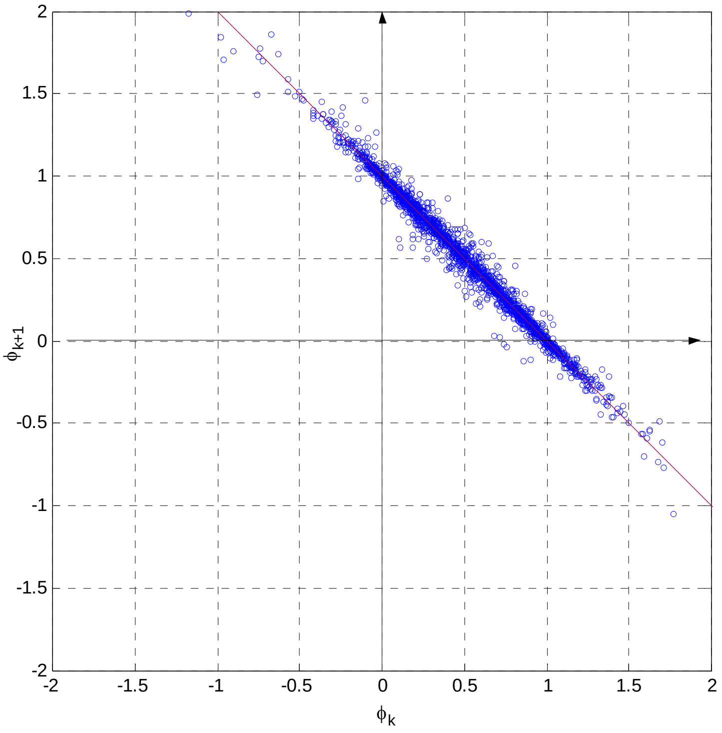
Figure 4-7. RTD with optimal (non-constrained) event functions. (1). Unconstrained event functions for RTD with given segmentation (i.e. event centers). Event centers are shown by circles. (2). LSF approximation, obtained by RTD decomposition. Original LSF trajectories are plotted by solid lines, while modeled ones are plotted by dashed lines. Event centers are shown by circles.

לפתרון (4.20) אפשר לתת פירוש גאומטרי במרחב וקטורי p-מימדי (כאשר p הוא אורך וקטורי הפרמטרים). בהינתן שני וקטורי המטרה השכנים, $\mathbf{a}_k, \mathbf{a}_{k+1}$, מבצעים קירוב של הוקטורים שבינהם $(\mathbf{y}(n), n_k \le n < n_{k+1})$ ע"י הטלתם על המישור שנפרש ע"י זוג וקטורי המטרה.

ניתן לראות (איור 4-ח), כי בפועל הטלות אלה אינן מכסות את המישור, אלא מתרכזות סביב הישר:

$$\hat{\mathbf{y}}(n) = \mathbf{a}_k \phi_k(n) + \mathbf{a}_{k+1}(1 - \phi_k(n)) \qquad (4.22)$$

המשמעות הגיאומטרית של הדבר, שההטלות על המישור הנפרש ע״י וקטורי המטרה, נופלות בפועל קרוב לישר המחבר בין שני וקטורים אלה לכן, מטעמי צמצום פרמטרי המודל לטובת הקוונטיזציה, ניתן לקרב את וקטורי הפרמטרים ע״י הטלה על הישר המחבר בין זוג וקטורי המטרה שמשני צידיהם.



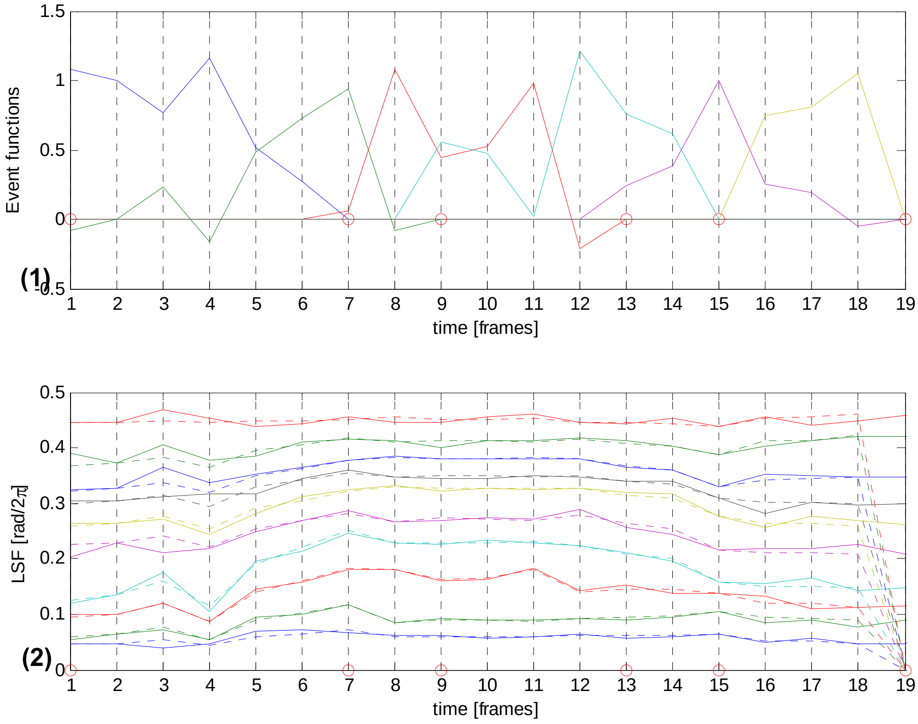
איור 4-ח. . פיזור של פונקציות מאורע רגעיות אופטימליות עבור מודל ה-RTD. פונקציות המאורע אשר מתקבלות לאחר ביצוע RTD עבור רצף של 2000 וקטורי LSF (מסדר 10). ניתן לראות, כי מרבית הנקודות נופלות בקרבת הקו הישר $\phi_k+\phi_{k+1}=1$ (המסומנת בקו רציף באיור)

Figure 4-8. Optimal instant event function scatter for RTD model. Shown are instant event function pairs, obtained by RTD, performed on 2000 LSF vectors (of 10th order). It may be easily seen that the scatter points fall close to $\phi_k+\phi_{k+1}=1$ straight line (drawn as a thin solid line).

במקרה זה הצבה של (4.22) לתוך (4.19) ומינימזציה ע״י גזירה לפי $\phi_k(n)$ והשוואה לאפס מביאה לפתרון:

$$\begin{pmatrix}\bar{\phi}_k(n)\\ \bar{\phi}_{k+1}(n)\end{pmatrix}=\begin{pmatrix}\dfrac{\left(\mathbf{y}(n)-\mathbf{a}_{k+1}\right)^T\left(\mathbf{a}_k-\mathbf{a}_{k+1}\right)}{\left(\mathbf{a}_k-\mathbf{a}_{k+1}\right)^T\left(\mathbf{a}_k-\mathbf{a}_{k+1}\right)}\\ 1-\bar{\phi}_k(n)\end{pmatrix},\ n_k\le n<n_{k+1} \qquad (4.23)$$

באיור 4-ט מוצגות פונקציות מאורע המשלימות ל-1 באזורים חופפים, בהתאם ל-(4.23). ניתן להיווכח, שעדיין צורות הפונקציות אינן רגולריות וקשות לקוונטיזציה, אבל בכל רגע ורגע מספיק לקודד רק פונקציית מאורע אחת, מתוך זוג המאורעות החופפים



איור 4-ט. . פירוק RTD עם פונקציות מאורע המשלימות ל-1. (1). פונקציות מאורעות המשלימות ל-1 באזורים חופפים עבור פירוק RTD עם סגמנטציה נתונה (כלומר מרכזי מאורעות ידועים). מרכזי המאורעות מסומנים בעיגולים. (1). קירוב ל-LSF, אשר התקבל ע״י פירוק ה-RTD המאויר ב-(1). ה-LSF המקוריים מסומנים בקווים רציפים, וה-LSF הממודלים מסומנים בקווים מקווקוים. מרכזי המאורעות מסומנים בעיגולים.

Figure 4-9. RTD with one's complementary event functions. (1). 1's complementary event functions for RTD with given segmentation (i.e. event centers). Event centers are shown by circles. (2). LSF approximation, gained by RTD decomposition, plotted in (1). Original LSF trajectories are plotted by solid lines, while modeled ones are plotted by dashed lines. Event centers are shown by circles.

לאחר השלמת מציאת פונקציות המאורע (בהתאם ל- (5.16)), ניתן לעדן את וקטורי המאורע ע״י מינימיזציה של (4.14) כפי שמתואר בסעיף 4.2.2. לאחר העידון, מתקנים מקדמי ה-LSF כך תישמר תכונת הסדר שלהם והפרש בין מקדמי LSF עוקבים (במסגרת מסוימת) לא יעלה על ε מסוים.

4.5 פירוק TD אופטימלי מבוסס RTD (ORTD)

4.5.1 תיאור כללי

ההרחבה הישירה לשיטת ה-RTD היא שיטת פירוק TD המכונה Optimized TD המוצעת ע״י Athaudage et al [44,450]. בעבודה זו אנו נכנה שיטה זו Optimized RTD או ORTD. בשיטה זו מנצלים את הפשטות בחישוב פונקציות המאורע בשיטת RTD (בהינתן מיקומי מרכזי המאורעות) ומשפרים את ביצועי האלגוריתם באמצעות אופטימיזציה על פני כל מיקום אפשרי של מרכזי המאורעות לבלוק אנליזה נתון, בהינתן מספר קבוע של מאורעות לבלוק M.

מודל ה-ORTD מתאר שיטה לביצוע ה-RTD בצורה ״המיטבית״ על פני *בלוק* אנליזה באורך N, בו מאלצים מספר קבוע של מאורעות M. (*בלוק* הוא אוסף וקטורי הפרמטרים הספקטרליים, עליהם מתבצעת פעולת הסגמנטציה תוך מזעור השגיאה הכוללת.) הדבר נעשה באמצעות חיפוש מלא על פני כל ה*סגמנטציות* האפשריות (כלומר מיקומי המאורעות האפשריים) ובחירת מיקומי המאורעות המשיגים את השגיאה הריבועית הכוללת המינימלית. בתהליך החיפוש מניחים, בדומה ל-RTD המקורי, את ההנחה הבאה לגבי ערכי וקטורי המטרה:

$$\mathbf{a}_k = \mathbf{y}(n_k) \tag{4.24}$$

במודל ה-RTD , כאמור, ניתן לחשב ערך פונקציות המאורע וכן שגיאת המודל בכל רגע n ע״י מינימיזציה של (4.19). Athaudage et al אינם מטילים אילוצים נוספים על פונקציות המאורע ולכן הפתרון האופטימלי (אשר מתקבל ע״י גזירה והשוואה לאפס של הנגזרת של (4.19)) יהיה:

$$\begin{pmatrix} \phi_k(n) \\ \phi_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_k^T\mathbf{a}_k & \mathbf{a}_k^T\mathbf{a}_{k+1} \\ \mathbf{a}_k^T\mathbf{a}_{k+1} & \mathbf{a}_{k+1}^T\mathbf{a}_{k+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_k^T\mathbf{y}(n) \\ \mathbf{a}_{k+1}^T\mathbf{y}(n) \end{pmatrix}, \quad n_k \le n \le n_{k+1} \tag{4.25}$$

כלומר ערכי פונקציות המאורעות (וכן שגיאת המודל הרגעית $E(n)$) תלויים רק ב-שני וקטורי המאורע השכנים. ההנחה ההתחלתית של RTD על וקטורי המאורע (4.21) תקפה גם כאן ולכן ניתן לרשום ערור שגיאת *הסגמנט* (n_k, n_{k+1}) ושגיאת *הבלוק* כולו (בלוק הוא אוסף וקטורי הפרמטרים הספקטרליים, עליו מתבצעת פעולת הסגמנטציה):

$$\begin{cases} E_{seg}(n_k, n_{k+1}) = \sum_{n=n_k}^{n_{k+1}-1} E(n) \\ E_{block}\left(n_0 \triangleq 0, n_1, \ldots, n_{M+1} \triangleq N+1\right) = \sum_{k=0}^{M} E_{seg}(n_k, n_{k+1}) \end{cases} \tag{4.26}$$

מאורע האפס ומאורע ה-$N+1$ הם מאורעות מדומים שיכולים להיות וקטורי אפס אם לא נאמר אחרת. נגדיר אלגוריתם תכנות דינמי למציאת סגמנטציה אופטימלית, במובן של מינימום שגיאה ריבועית ממוצעת (כלומר שגיאה מצטברת מינימלית של הבלוק כולו), ע״י:

$$\begin{cases} D(n_k) = \min\limits_{n_{k-1}\in R_{k-1}} \left(D(n_{k-1}) + E_{seg}(n_{k-1}, n_k)\right) \\ n_{k-1}^{*} = \arg\min\limits_{n_{k-1}\in R_{k-1}} \left(D(n_{k-1}) + E_{seg}(n_{k-1}, n_k)\right) \end{cases}, \; k = 2, \ldots M \tag{4.27}$$

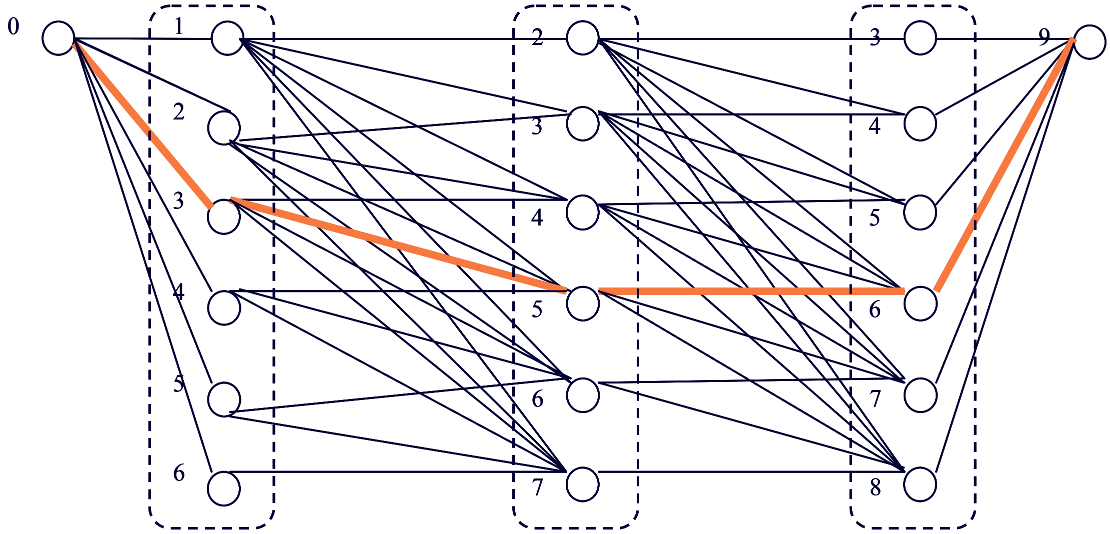
כאשר $D(n_k)$ הינה השגיאה המצטברת עד לרגע n_k, המוגדרת ע״י:

$$D(n_k) = \sum_{n=1}^{n_k - 1} E(n) \tag{4.28}$$

ו- R_k הוא טווח החיפוש, אשר יוגדר לאור חיפוש מלא ע״י:

$$R_k = \{ n \mid n_{k-1} < n < n_{k+1} \} \tag{4.29}$$

ניתן לישם את החיפוש המלא בצורה יעילה באמצעות חיפוש סבכה (trellis) עפ״י אלגוריתם Viterbi [2]. דוגמה לשבכת חיפוש ניתנת באיור 4-י.



איור 4-י. דוגמה לדיאגרמת trellis המשמשת לחיפוש אתר הסגמנטציה האופטימלית ב-ORTD. כל צומת מציינת קצה סגמנט אפשרי, השלבים (המלבנים המקווקוים) מאחדים את כל המיקומים האפשריים של המאורע מסוים והקשתות מציינות את הסגמנטים החוקיים. הדוגמה ניתנת עבור N=8 ו-M=3. מאורע ה-אפס וה-N+1 מציינים את הקצוות של הבלוק (מאורעות מדומים). הקו העבה מציין את הסגמנטציה האופטימלית

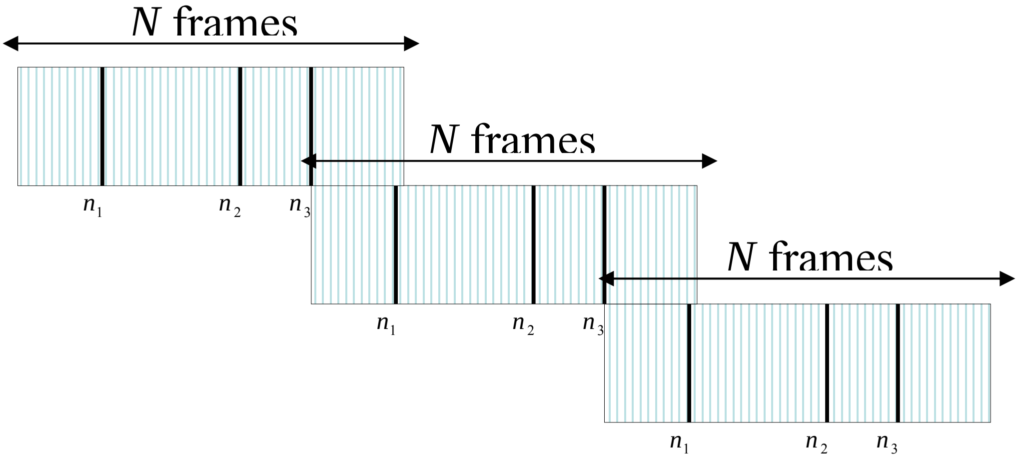
Figure 4-10. . A trellis example for best segmentation search in ORTD. Each node designates a possible segment edge, the stages (dashed rectangles) point out all the possible placements of a specific event and the arcs define the allowable segments. Here N=8 and M=3. The zero and N+1-th events are dummy events, indicating the block edges. The thick line indicates an optimal segmentation.

לאחר השלמת מציאת פונקציות המאורע האופטימליות, יש לעדן את וקטורי המאורע ע״י מינימיזציה של (4.14) בהתאם לסעיף 4.2.2.

ניתן לבצע מספר איטרציות של האלגוריתם כולו (חישוב פונקציות המאורע ועידון וקטורי המטרה), כאשר באיטרציות הבאות שינוי מרכזי המאורעות כבר לא משפיע על ערך וקטורי המאורע (שהתקבלו לאחר איטרציה קודמת) ולכן יש לבצע פחות חישובים ע״מ לחשב כל שגיאות הסגמנטים האפשריים לצורך ביצוע האופטימיזציה. המחברים טוענים שאין צורך ביותר מ-5 איטרציות כיוון שהשיפור נהיה זניח לאחר מכן [45]. נשים לב, כי ההפעלה הראשונה של שלב מציאת פונקציות המאורעות (והסגמנטציה האופטימלית) מכילה בתוכה גם שלב של קביעה ראשונית של וקטורי המטרה (ע״י $\mathbf{a}_k = \mathbf{y}(n_k)$) וקובעת למעשה את נקודת העבודה של המודל, בהתאם להנחות ה-RTD והנחת הקשר הישיר בין וקטור המטרה לבין מיקום מרכז המאורע. השלבים הבאים יבצעו עידון הפרמטרים לאחר שלב התחלתי זה, והשיפור משלב לשלב מובטח. אמנם הפתרון אליו האלגוריתם יתכנס הוא רק <u>מינימום לוקלי</u> בסביבת נקודת העבודה שנקבעה בשלב ההתחלתי, למרות החיפוש המלא של כל הסגמטציות האפשריות. הדבר נובע מכך, שהחיפוש המלא של הסגמנטציות בשלב הראשוני מסתמך על ההנחה, שנכונה רק בקירוב.

4.5.2 חפיפה בין הבלוקים ותנאי קצה

ביצועים טובים של אלגוריתם זה מותנים בחפיפה בין בלוקים עוקבים, כיוון שהפתרון האופטימלי המוצע, עם אירועי אפס מדומים בהתחלה ובסוף, אינו ממדל היטב את קצות הבלוק. גודל החפיפה בין הבלוקים העוקבים שנבחר ב-ORTD הוא כגודל הסגמנט האחרון בבלוק, כלומר וקטור המאורע האחרון של הבלוק הקודם הוא מאורע האפס של הבלוק הנוכחי (ר׳ איור 4-יא).



איור 4-יא. החפיפה בין החוצצים ב-ORTD.

Figure 4-11. Buffer overlap in ORTD.

למרות החפיפה, מספר המסגרות שבבלוק נשמר שווה ל-N, כלומר קצב המאורעות באלגוריתם ה-ORTD הנו קצב משתנה הגבוה במקצת מ- $f_{frm}\dfrac{M}{N}$, כאשר f_{frm} הוא קצב המסגרות לשנייה, N אורך הבלוק ו-M מספר מאורעות בבלוק. כיוון שהמאורע המדומה שמספרו $M+1$ נשאר עדיין מאורע האפס, הפתרון האופטימלי ימקם מאורע אחרון קרוב לקצה הבלוק, ולכן אורך החפיפה לא יהיה גדול מדי.

המחברים מפעילים בהצלחה את האלגוריתם על מקדמי ה-LSF למרות שאין מתקיימת כאן בהכרח *תכונת הסדר* של מקדמי ה-LSF (ניתן לאלץ אותה לאחר החישוב ע"ח הגדלה קלה בשגיאת המודל).

אלגוריתם ה-ORTD כהגדרתו לא נועד לשימוש בסביבת זמן אמת (כנדרש בד"כ ערור מקודדי דיבור) עקב השימוש בפתרון של החיפוש המלא. בפרק הבא נעריך את סיבוכיות האלגוריתם ORTD ונציג את אלגוריתם ההמשך.

4.6 סיכום

הצגנו, אפוא, שיטות שונות ליישום של מודל ה-TD. התחלנו מתיאור השיטה המקורית של Atal והמשכנו עם הגישות המודרניות המותאמות לקידוד מקדמי ה-LSF. תיארנו תחילה שתי שיטות לחישוב פונקציות המאורע, בהינתן הסגמנטציה ווקטורי המטרה ההתחלתיים. החיסרון של השיטות הנ"ל בכך שהן מסתמכות על הקביעה ההתחלתית של מרכזי המאורעות, אשר בד"כ אינה אופטימלית, ולכן יש צורך להשתמש במספר רב (יחסית למספר הפונמות האמיתיות) של מאורעות ע"מ לתאר נאות את ספקטרום הדיבור (מדובר בכ-18-20 מאורעות לשנייה בממוצע [41,40]), דבר שמעלה את מספר הסיביות שיש להשקיע ע"מ לקודד את המעטפת הספקטרלית. אחת השיטות הללו, הקרויה RTD, מניבה פתרון אנליטי פשוט לבעיית מציאת פונקציות המאורעות, דבר שמאפשר לבצע חיפוש על פני כל הסגמנטציות האפשריות ולבחור זו, שמביאה ל-RTD עם השגיאה הכוללת המינימלית (על פני *הבלוק* של מסגרות). זאת, למעשה, השיטה האחרונה שתוארה בפרק זה, המכונה Optimized RTD (ORTD). בפרק הבא נרחיב ונייעל את אלגוריתם ה-ORTD ע"מ להקטין את שגיאת ההתאמה של ה-ORTD ולהקטין את העומס החישובי שלו.

פרק 5

קידוד מעטפת דיבור באמצעות
Dynamically Weighted Sub-Optimal RTD (DW-SORTeD)

5.1 מבוא

בפרק זה נתאר אלגוריתם פירוק TD חדש, שמוצע בעבודה זו ואשר מתבסס על אלגוריתם ה-ORTD המתואר לעיל אך מותאם לצרכי קידוד דיבור בקצבים נמוכים ביותר בזמן אמת. להלן השינויים שהוכנסו באלגוריתם ה-ORTD:

1) הכללה של ORTD עבור קריטריון של שגיאה ריבועית משוקללת (דוגמת G-WSE או PA-WSE) לשיפור האיכות התפיסתית (perceptual) של ההתאמה הספקטרלית

2) פיתוח אלגוריתם תת-אופטימלי יעיל (סיבוכיות של $O(N)$ במקום $O(N^2)$ למסגרת) עבור מציאת פונקציות המאורע, אשר מחליף את החיפוש המלא של ה-ORTD.

3) המרה של פתרון אופטימלי לפונקציות המאורע (בהינתן וקטורי ומיקומי המאורעות) (4.25) בפתרונות מאולצים לטובת השיפור בקוונטיזציה, דוגמת (5.16), (5.19) ואילוצים נוספים לצורך פישוט הקוונטיזציה.

מבנה הפרק הוא כדלהלן. ראשית, נסביר את הצורך באלגוריתם החדש. לאחר מכן נתאר הרחבה ישירה של אלגוריתם ה-ORTD הקרויה Dynamically Weighted Optimized RTD (DW-ORTD). בסעיפים הבאים לאחר מכן נציע מודיפיקציה תת-אופטימלית של ה-ORTD, המכונה Sub-Optimal RTD (SORTeD). המודל הסופי שמשלב מתוכו שגיאה ריבועית ממשוקללת וחיפוש תת-אופטימלי יכונה Dynamically Weighted Sub-Optimal RTD (DW-SORTeD). לבסוף, נתאר את מערכת ה-TD בשילוב עם קוונטיזציה של מקדמי ה-LSF.

5.2 בחירת הקריטריונים לקירוב המעטפת הספקטרלית

5.2.1 כללי

רוב האלגוריתמים לפירוק TD משתמשים בקריטריון של שגיאה ריבועית ממוצעת למציאת פונקציות ווקטורי המאורעות. ידוע כי השגיאה הריבועית אינה מתארת בהכרח בצורה נאותה את מידת המרחק כפי שהיא *נתפסת* ע״י מערכת השמיעה האנושית. למשל, האוזן רגישה יותר לשגיאות באזורי הפורמנטים (כלומר לאזורים של השיאים המקומיים של המעטפת), מאשר באזורים של ספקטרום בעל ערכים נמוכים. קריטריון מרחק תפיסתי (perceptual) מקובל הוא LSD (ר׳ סעיף 2.4), אך לא ניתן לשלבו בתוך ה-ORTD ו/או בקוונטיזציה, בעיקר עקב הסיבוכיות הגבוהה שלו. קיימים אכן מדדי מרחק, המתקרבים ל-LSD ו/או ממדלים את מידת הקרבה התפיסתית. המדדים המקובלים (בעולם הקידוד) מתבססים על שגיאה ריבועית משוקללת, עם משקלות משתנים (התלויים בווקטור הפרמטרים המקורי), למשל המדדים G-WSE ו-PA-WSE (ר׳ סעיף 2.4) ועוד. באמצעות מדדים אלה ניתן לאמוד ביתר נאמנות את המרחק התפיסתי בין הפרמטרים הספקטרליים הממודלים (או המקוונטים) למקוריים. אנחנו נצפה לקבל ערכי שגיאת LSD במוצא המערכת, שממזערת שגיאות משוקללות אילו, שהם נמוכים מאילו המתקבלים ממזעור שגיאה ריבועית ללא שקלול.

5.2.2 מדד ה-G-WSE המסונן

בחינת ביצועים של מדדי שגיאה ריבועיים לצרכי ה-TD הביאה להגדרה של מדד נוסף, המתבסס על מדד ה- G-WSE. הוא נוצר ע״י הורדת החשיבות של מקדמי ה-LSF העליונים, אשר מאפיינים את החלק העליון של ספקטרום הדיבור (כלומר סינון מעביר נמוכים):

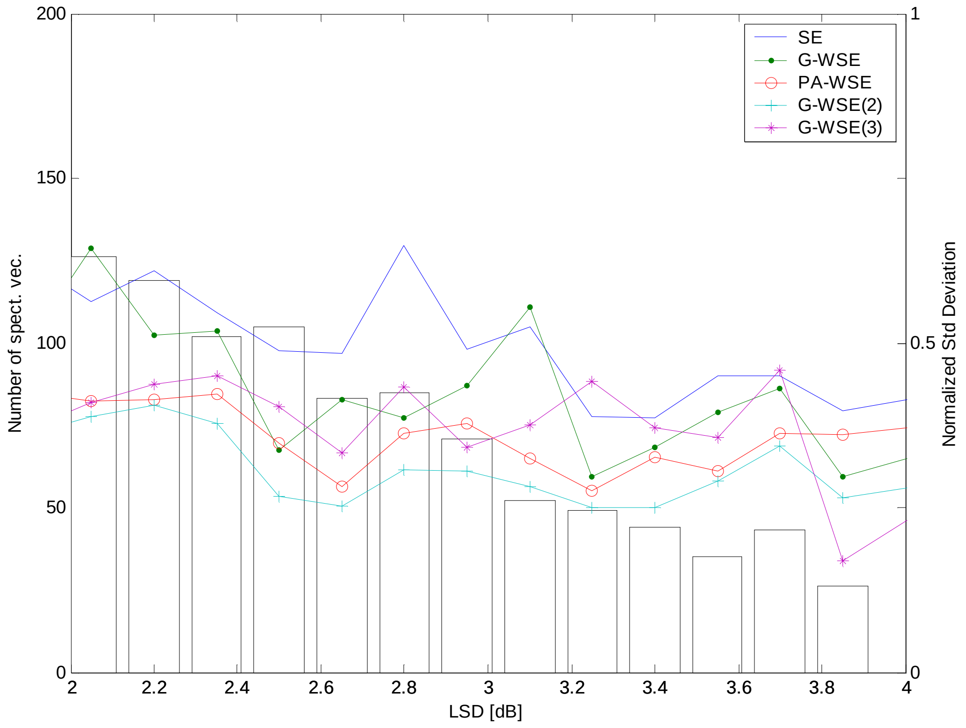
$$\begin{aligned}\tilde{w}_i(n) &= (c_i)^2 w_i(n),\\ \mathbf{c} &= [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0.9 \quad 0.8 \quad 0.7 \quad 0.1 \quad 0.01]\end{aligned} \qquad (5.1)$$

כאשר $w_i(n)$ זהו משקל Gardner של הרכיב ה-i ברגע זמן n (ר׳ סעיף 2.4).

מדד זה, שיסומן בהמשך G-WSE(2), נמצא ונותן התאמה משופרת למדד ה-LSD, המוערכת ע״י בחינת סטיית התקן המנורמלת של ערכי מדד זה על פני ה-histogram bins של ה-LSD (ר׳ איור 5-א). מדד זה הוכיח את עצמו גם בבדיקות השוואתיות מסכמות (ר׳ פרק 6).

ברור כי הורדה כה קיצונית של חשיבות התדרים הגבוהים הייתה פוגעת באיכות הדיבור המשוחזר אילו מדד זה היה משמש ישירות לקוונטיזציה של וקטורי ה-LSF. אולם מדד זה משמש בהצלחה כקריטריון למינימיזציה בפירוק ה-TD, בהינתן ייצוג הולם של וקטורי המטרה, כיוון שהוא מדגיש אזורים ספקטרליים שחשובים יותר לאוזן האנושית. אלה מכתיבים באמצעות פונקציות המאורע את דרך ההשתנות הזמנית של הספקטרום. באיור 5-א מוצגת בנוסף שגיאת G-WSE המסוננת סינון עמוק יותר (הירידה מתחילה מהרכיב הרביעי). שגיאה זו מסומנת כ-G-WSE(3). ההתאמה של

ה-G-WSE(3) ל-LSD גרועה מזו של G-WSE(2) ושל שגיאות נוספות. הסיבה לכך היא בכך שהנחתה של אזור מרכזי אשר מכיל פורמנטים חזקים מדרדרת את יכולת ההתאמה לשגיאת ה-LSD.



איור 5-א. יכולת התאמה למדד ה-LSD של שגיאות ריבועיות משוקללות. סטיית התקן של שגיאות ריבועיות, המנורמלת ע"י הממוצע שלהן. מחושב על פני histogram bins של LSD, המתוארים גם הם באיור זה.ככל שהסטייה נמוכה יותר, כך ההתאמה טובה יותר. הסימונים במקרא הם כדלהלן: SE – שגיאה ריבועית, G-WSE – שגיאה ריבועית משוקללת של Gardner, PA-WSE – שגיאה ריבועית משוקללת של Paliwal-Atal, G-WSE(2) - שגיאה ריבועית משוקללת של Gardner, מסוננת החל המרכיב השישי בהתאם ל-(5.1), G-WSE(3) - שגיאה ריבועית משוקללת של Gardner, מסוננת החל המרכיב הרביעי.

Figure 5-1. LSD matching ability of weighted squared errors. Standard deviation of squared errors, normalized by their averages, calculated over LSD histogram bins. The lower the deviation is, the better matching ability is obtained. The histogram of the utterance LSD is also presented. The legend notation is as follows: SE – Squared Error, G-WSE – Gardner Weighted SE, PA-WSE – Paliwal-Atal Weighted SE, G-WSE(2) – LP filtered Gardner WSE, starting from 6th component according to (5.1), G-WSE(3) – LP filtered Gardner WSE, starting from 4th component.

5.3 פירוק RTD אופטימלי (ORTD) עם קריטריון של WSE דינמי (DW-ORTD)

בסעיף זה נציג הרחבה ישירה של האלגוריתם ORTD (ר' סעיף 4.5), המאפשרת מינימיזציה לפי קריטריון שגיאה ריבועית משוקללת עם משקלים דינמיים. למודיפיקציה זו נקרא Dynamically Weighted ORTD (DW-ORTD).

5.3.1 מודיפיקציות לשלב של מציאת פונקציות המאורעות

השגיאה שיש למזער (עבור כל בלוק של מסגרות באורך N המיוצג ע"י מטריצת הפרמטרים הספקטרליים $\mathbf{Y}$ בגודל $p \times N$) הינה:

$$(5.2)\; E(n) = \left(\mathbf{y}(n) - \hat{\mathbf{y}}(n)\right)^T \mathbf{W}(n)\left(\mathbf{y}(n) - \hat{\mathbf{y}}(n)\right), \qquad \mathbf{W}(n) = \begin{pmatrix} w_1(n) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_p(n) \end{pmatrix}$$

כאשר $\hat{\mathbf{y}}(n)$ הינו וקטור המשוערך ע"י מודל ה-RTD ונתון ע"י (4.18) ו-$\mathbf{W}(n)$ היא מטריצה אלכסונית של המשקלים התלויים בזמן n. מזעור של (5.2) ביחס לפונקצית המאורע נותן (בהינתן וקטורי המאורע ומיקומי מרכזי המאורעות):

$$(5.3) \qquad \begin{pmatrix} \phi_k(n) \\ \phi_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_k^T \mathbf{W}(n)\mathbf{a}_k & \mathbf{a}_k^T \mathbf{W}(n)\mathbf{a}_{k+1} \\ \mathbf{a}_k^T \mathbf{W}(n)\mathbf{a}_{k+1} & \mathbf{a}_{k+1}^T \mathbf{W}(n)\mathbf{a}_{k+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_k^T \mathbf{W}(n)\mathbf{y}(n) \\ \mathbf{a}_{k+1}^T \mathbf{W}(n)\mathbf{y}(n) \end{pmatrix},$$

האלגוריתם הדינמי למציאת פונקציות המאורע האופטימליות אינו שונה מהאלגוריתם המקורי, יש רק להמיר את קריטריון השגיאה הריבועית בקריטריון (5.2) ולהשתמש ב-(5.3) לצרכי החישוב של פונקציות מאורע רגעיות.

5.3.2 מודיפיקציות לשלב של עידון וקטורי המטרה

לעומת זאת, עידון וקטורי המטרה הוא שונה, מכיוון שהמשקלים ב-WSE אינם קבועים בזמן. לאחר השלמת מציאת הסגמנטציה האופטימלית (יחד עם פונקציות מאורע אופטימליות, בהתאם ל-(5.3)) , ניתן לעדן את וקטורי המאורע ע"י מינימיזציה של שגיאת הבלוק הכוללת כפונקציה של וקטורי המטרה. בשלב זה של האלגוריתם יש למצוא את מטריצת וקטורי המטרה הממזערת את מדד ה-WMSE לבלוק, בהינתן פונקציות המאורעות $\phi_{i,n} \triangleq \phi_i(n)$ ומיקומי המאורעות n_i , ז"א יש למזער את:

(5.4)

$$E_{block}(\mathbf{A}) = \sum_k \sum_{n=n_k}^{n_{k+1}-1} = (\mathbf{y}(n) - \mathbf{a}_k\phi_k(n) - \mathbf{a}_{k+1}\phi_{k+1}(n))^T \mathbf{W}(n)(\mathbf{y}(n) - \mathbf{a}_k\phi_k(n) - \mathbf{a}_{k+1}\phi_{k+1}(n))$$

גזירה של הביטוי לפי וקטורי המטרה המתחשב בתכונות ה-RTD נותנת (נספח I) :

$$\begin{pmatrix} \mathbf{D}_1 & \mathbf{X}_1 & 0 & 0 \\ \mathbf{X}_1 & \ddots & \ddots & 0 \\ 0 & \ddots & \mathbf{D}_{M-1} & \mathbf{X}_{M-1} \\ 0 & 0 & \mathbf{X}_{M-1} & \mathbf{D}_M \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{M-1} \\ \mathbf{a}_M \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{M-1} \\ \mathbf{b}_M \end{pmatrix}, \tag{5.5}$$

כאשר $\mathbf{X}_k = \sum_n \phi_k(n)\phi_{k+1}(n)\mathbf{W}(n)$, $\mathbf{D}_k = \sum_n \phi_k^2(n)\mathbf{W}(n)$ ו- $\mathbf{b}_k = \sum_n \phi_k(n)\mathbf{W}(n)\mathbf{y}(n)$.

ו- $\mathbf{a}_i$ הינו וקטור המטרה ה-i.

כיוון שכל תת-המטריצות המרכיבות את המטריצה הגדולה ב-(5.5) הן אלכסוניות, ניתן לפרק את המשואה המטריצית (5.5) ל-p מערכות משואות לינאריות (כמספר הרכיבים בוקטורי הפרמטרים) :

$$\begin{pmatrix} d_{i,1} & x_{i,1} & 0 & 0 \\ x_{i,1} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,M-1} & x_{i,M-1} \\ 0 & 0 & x_{i,M-1} & d_{i,M} \end{pmatrix} \begin{pmatrix} a_{i,1} \\ \vdots \\ a_{i,M-1} \\ a_{i,M} \end{pmatrix} = \begin{pmatrix} b_{i,1} \\ \vdots \\ b_{i,M-1} \\ b_{i,M} \end{pmatrix}, \quad 1 \le i \le p, \tag{5.6}$$

כאשר $x_{i,k} = \sum_n \phi_k(n)\phi_{k+1}(n)w_i(n)$, $d_{i,k} = \sum_n \phi_k^2(n)w_i(n)$ ו- $b_{i,k} = \sum_n \phi_k(n)w_i(n)y_i(n)$.

משואות אילו הן תלת-אלכסוניות וניתנות לפתרון ע"י תהליך חילוץ גאוס (Gauss elimination) , כפי שהוסבר בסעיף 4.2.2.

אם $x_{i,k}$ הוא קרוב לאפס (עקב חוסר קורלציה בין פ' המאורע השכנות – דבר שעלול לקרות אם מרכזי מאורעות עוקבים שוכנים זה לצד זה) המטריצה הינה סינגולרית (או קרובה לכך). במקרה זה המטריצה היא בלוק-אלכסונית, וניתן לפרק מערכות את המשוואות הלינאריות ב-(5.6) למספר תת-בעיות קטנות, בהתאם לחלוקה ב-(5.7) למטריצות בעלות דרגה מלאה.

$$\begin{bmatrix} \begin{bmatrix} d_{i,1} & x_{i,1} & 0 & 0 \\ x_{i,1} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,m-1} & x_{i,m-1} \\ 0 & 0 & x_{i,m-1} & d_{i,m} \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} d_{i,m+1} & x_{i,m+1} & 0 & 0 \\ x_{i,m+1} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,M-1} & x_{i,M-1} \\ 0 & 0 & x_{i,M-1} & d_{i,M} \end{bmatrix} \end{bmatrix} \tag{5.7}$$

כאמור, הבלוקים של וקטורי הפרמטרים בהם מתבצעת האופטימיזציה הם חופפים, ז״א וקטור המטרה ה-0 של הבלוק הנוכחי הינו וקטור המטרה האחרון של הבלוק הקודם, לכן אין לשנות אותו, אך יש לקחתו בחשבון. בהתחשב בוקטור המטרה ה-0, מערכות המשוואות (5.6) ייראו:

$$\begin{pmatrix} d_{i,1} & x_{i,1} & 0 & 0 \\ x_{i,1} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,M-1} & x_{i,M-1} \\ 0 & 0 & x_{i,M-1} & d_{i,M} \end{pmatrix} \begin{pmatrix} a_{i,1} \\ \vdots \\ a_{i,M-1} \\ a_{i,M} \end{pmatrix} = \begin{pmatrix} b_{i,1} - x_{i,0}a_{i,0} \\ \vdots \\ b_{i,M-1} \\ b_{i,M} \end{pmatrix}, \quad 1 \le i \le p. \tag{5.8}$$

כזכור, פירוק ה-ORTD מופעל על בלוקים חופפים, והמאורע האחרון של כל בלוק הוא למעשה המאורע הראשון של הבלוק הבא. עקב כך, השינוי של וקטור המטרה האחרון (כחלק משלב עדון וקטורי המטרה בבלוק הנוכחי) עלול להביא לקלקול בבלוק הבא. ראינו בניסויים, שלא כדאי בחלק מהמקרים (כאשר משתמשים בפונקציות המאורעות מאולצות או מקוונטות) לעדן את המאורע האחרון בשלב של עידון וקטורי המטרה (שיפור של 0.07 dB נצפה כתוצאה מביטול העידון של וקטור המטרה האחרון, כפי שניתן לראות בסעיף 6.2.50). במקרה זה המשוואה (5.8) הופכת ל-

$$\begin{pmatrix} d_{i,1} & x_{i,1} & 0 & 0 \\ x_{i,1} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,M-2} & x_{i,M-2} \\ 0 & 0 & x_{i,M-2} & d_{i,M-1} \end{pmatrix} \begin{pmatrix} a_{i,1} \\ \vdots \\ a_{i,M-2} \\ a_{i,M-1} \end{pmatrix} = \begin{pmatrix} b_{i,1} - x_{i,0}a_{i,0} \\ \vdots \\ b_{i,M-2} \\ b_{i,M-1} - x_{i,M-1}a_{i,M} \end{pmatrix}, \quad 1 \le i \le p. \tag{5.9}$$

הצגנו, אפוא, הרחבה ישירה של אלגוריתם ה-ORTD הקרויה DW-ORTD, אשר מאפשרת פרוק RTD אופטימלי במובן מזעור של שגיאה ריבועית משוקללת התלויה בזמן. האלגוריתם החדש הוא

בעל בסיבוכיות הדומה לזו של האלגוריתם המקורי והוא מאפשר עקיבה משופרת אחר ההשתנויות של שגיאת ה-LSD אם הוא מופעל על שגיאות ריבועיות משוקללות מתאימות.

5.4 אלגוריתם תת-אופטימלי למציאת פונקציות המאורע.

5.4.1 מבוא

האלגוריתם האופטימלי ORTD וההרחבה הישירה שלו DW-ORTD , המתוארים לעיל, מבצעים חיפוש מלא של כל המיקומים האפשריים של M מרכזי מאורעות בתוך הבלוק הנתון, למציאת המיקומים שמביאים למינימום את השגיאה הכוללת של הבלוק. הדבר ניתן ליישום, כיוון שבמודל ה-RTD שגיאת המודל של סגמנט בתוך הבלוק, אשר נמצא בין מרכזים של שני מאורעות עוקבים, תלויה רק במיקום של מרכזי מאורעות אלו ובערכי וקטורי המטרה שלהם :

$$E_{seg}(n_k, n_{k+1}) = \sum_{n=n_k}^{n_{k+1}-1} E(n, \mathbf{a}_k, \mathbf{a}_{k+1}), \qquad (5.10)$$

כאשר $E(n, \mathbf{a}_k, \mathbf{a}_{k+1})$ הנה שגיאת מודל רגעית הניתנת ע"י :

(5.11)

$$E(n, \mathbf{a}_k, \mathbf{a}_{k+1}) = \sum_n (\mathbf{y}(n) - \mathbf{a}_k\phi_k(n) - \mathbf{a}_{k+1}\phi_{k+1}(n))^T \mathbf{W}(n)(\mathbf{y}(n) - \mathbf{a}_k\phi_k(n) - \phi_{k+1}(n)),\ n_k \le n < n_{k+1},$$

$$\mathbf{W}(n) = \begin{pmatrix} w_1(n) & 0 & \mathbf{0} \\ 0 & \ddots & 0 \\ \mathbf{0} & 0 & w_p(n) \end{pmatrix}$$

וצמד פונקציות המאורע הרגעיות $\phi_k(n), \phi_{k+1}(n)$ מחושבות בצורה אופטימלית, כמו ב-(5.3) או ע"י פתרונות מאולצים למיניהם ((5.16), (5.19) ועוד).

ישנן מספר סיבות שהובילו לפיתוח אלגוריתם תת-אופטימלי (SORTeD) :

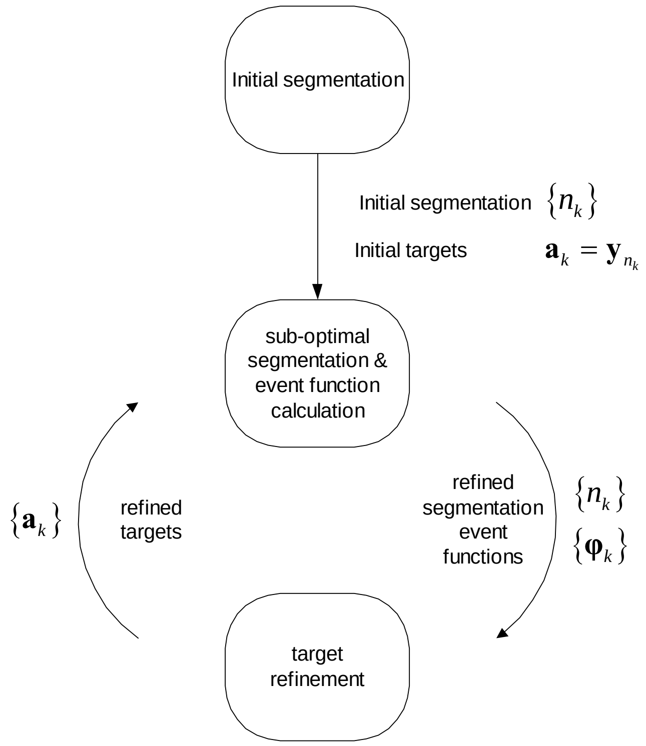
1) הסיבוכיות של החיפוש המלא שמופעל ב-ORTD גבוהה מכדי להשתמש בשיטה זו במקודדי דיבור בזמן אמת, ולכן מתעורר הצורך לפתח אלגוריתם חיפוש חלקי (תת-אופטימלי) למציאת פונקציות המאורע.

2) באלגוריתם ORTD המקורי קיימת חפיפה בין הבלוקים העוקבים, אך גודל כל בלוק נשאר קבוע, כלומר מספר מסגרות *חדשות* בבלוק עלול להשתנות. עקב כך למרות האילוץ של M מאורעות לבלוק, אלגוריתם ה-ORTD פועל בקצב מאורעות משתנה. קצב משתנה במקודד עלול לסבך ולייקר את מימושו בחמרה, לכן רצוי שיהיה למקודד קצב תמסורת קבוע.

האלגוריתם התת-אופטימלי הקרוי Sub-Optimal RTD (SORTeD) הנו אלגוריתם לביצוע ה-RTD עם קריטריון השגיאה הריבועית או השגיאה הריבועית המשוקללת. השם המלא לאלגוריתם התת-אופטימלי שמשתמש בקריטריון ה-WMSE עם משקלות דינמיים הנו Dynamically Weighted SORTeD (DW-SORTeD). כיוון שמשקלות דינמיים משפרים את הביצועים של האלגוריתם ללא הוספה משמעותית לעומס החישובי, מכאן והלאה אנחנו תמיד נשתמש בהם, אלא אם נאמר אחרת, ונכנה את האלגוריתם התת אופטימלי בשמו המקוצר, SORTeD. אלגוריתם זה פועל הן בקצב קבוע הן בקצב משתנה ומתכנס למינימום מקומי של קריטריון השגיאה המצטברת לבלוק נתון, בהתאם לתנאים התחלתיים של האלגוריתם, שהם מרכזי המאורעות $\left\{n_k\right\}_{k=1}^M$. ניתן לקרוא למידע זה סגמנטציה התחלתית, כיוון שמרכזים אלו גם קובעים את אורכי פונקציות המאורע ב-RTD. בסעיף זה יתואר תחילה האלגוריתם בקווים כלליים (ר' איור 5-ב) ולאחר מכן הוא יפורט.

5.4.2 תיאור כללי

התרשים הכללי של האלגוריתם ניתן באיור 5-ב. בשלב ראשון של האלגוריתם קובעים סגמנטציה ראשונית של הבלוק $\left\{n_k\right\}_{k=1}^M$. הסגמנטציה ההתחלתית מוזנת למודול שמוצא סגמנטציה משופרת (תת-אופטימלית) ופונקציות המאורע המתאימות לה. אלו מוזנים למודול שמחשב את וקטורי המטרה החדשים (אופטימליים), בהינתן פונקציות המאורע. וקטורי המטרה, לאחר עידונם (refinement), מוזנים חזרה למודול של חיפוש הסגמנטציה התת-אופטימלי. ניתן לבצע מספר איטרציות של האלגוריתם (בד"כ אחת או שתיים מספיקות). בפעם הראשונה שמודול הסגמנטציה מתבצע, הסגמנטציה $\left\{n_k\right\}_{k=1}^M$ קובעת גם את וקטורי המטרה ע"י $\mathbf{a}_k = \mathbf{y}_{n_k}$ לצורך חישוב של פונקציות המאורע ושגיאות מודל ה-RTD במהלך החיפוש. באיטרציות הבאות מוזנים אליו וקטורי המטרה המעודנים (refined) והם אלה שישמשו לחישוב פונקציות המאורע ופונקציות השגיאה במהלך החיפוש.



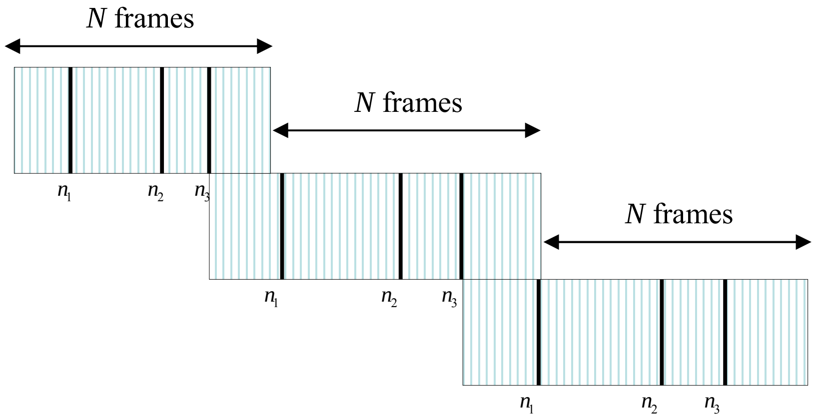
איור 5-ב. אלגוריתם ה-RTD התת-אופטימלי (SORTeD). תיאור כללי

Figure 5-2. Sub Optimal RTD (SORTeD) algorithm. General description

5.4.3 עדכון של הבלוק לאנליזה

תהי $\mathbf{Y}$ מטריצה המכילה $\tilde{N}$ וקטורים ספקטרליים, עליה מופעל אלגוריתם ה-SORTeD. נתייחס אליה מכאן ואילך כאל *בלוק האנליזה* הנוכחי. בדומה ל-ORTD, בלוק אנליזה נוכחי חופף בד"כ בחלקו עם בלוק אנליזה קודם, כך שתחילתו מתלכדת עם מיקום מרכז המאורע האחרון של הבלוק הקודם (גודל החפיפה לא קבוע). נסמן ב-N מספר המסגרות החדשות בתוך בלוק האנליזה (כלומר, ישנם $N-\tilde{N}$ מסגרות חופפות). נתייחס לשתי אפשרויות לעדכון של בלוק אנליזה:

1) <u>קצב עדכון קבוע (אורך הבלוק משתנה)</u> - $\tilde{N}$ אינו קבוע, אך מספר מסגרות חדשות בכל בלוק ($N<\tilde{N}$) הוא קבוע. במתכונת זו נוכל להגיע למקודד בעל קצב תמסורת קבוע (M מאורעות משודרות לכל- N מסגרות). זוהי השיטה המועדפת במקודד שלנו (ר' איור 5-ג).



איור 5-ג. חפיפה בין הבלוקים השומרת על קצב קבוע של המאורעות לשנייה.

Figure 5-3. Analysis block overlap for constant rate scheme.

2) <u>אורך בלוק קבוע (קצב עדכון משתנה)</u> – מספר מסגרות בבלוק $\tilde{N}$ הוא קבוע, אך מתוכם מספר לא קבוע של מסגרות חדשות (בהתאם לגודל החפיפה) בנוסף לאלו החופפות עם בלוק קודם. שיטה זו מניבה מקודד בעל קצב תמסורת משתנה אך כמות החישובים זהה בכל בלוק אנליזה. שיטה זו נהוגה ב-ORTD שמוגדר כאלגוריתם להפעלה offline (ר' איור 4-יא).

5.4.4 קביעה של סגמנטציה התחלתית

יהי M – מספר המאורעות חדשים שמעוניינים לזהות בבלוק האנליזה הנוכחי. הסגמנטציה הראשונית $\{n_k\}_{k=0}^{M+1}$ היא זו שמוזנת לתוך מודול החיפוש התת-אופטימלי, כאשר נתייחס למאורעות המדומים n_0 ו- n_{M+1} כאל *קצות הבלוק*. מאורעות אלה אינם משודרים אך מסייעים במציאת הסגמנטציה ומגדירים למעשה גבולות אפקטיביים של בלוק האנליזה הנוכחי שעבורו יש למזער את השגיאה המצטברת. כמובן שככל שהסגמנטציה הראשונית תהיה קרובה יותר לסגמנטציה המיטבית, כך הפתרון התת-אופטימלי יתקרב לפתרון המיטבי. באופן טבעי, ניתן לקבוע חלוקה ראשונית בהתחשב במדד ה-SFTR (4.15) (למשל ע״י בחירת נקודות מינימום מקומי), אך כפי שהתברר, אלגוריתם החיפוש אינו רגיש במיוחד לתנאי התחלה, פרט אולי למיקום המאורע האחרון. לכן נבחר חלוקת מאורעות ראשונית אחידה, למעט מאורע אחרון שייבחר בהתאם לקריטריון ה-SFTR או בקצה הבלוק.

כיוון שפנינו לקראת אלגוריתם מעשי, אין לאפשר חפיפה גדולה מדי מטעמי סיבוכיות. חפיפה לא מוגבלת יכולה להביא לניפוח רב של דרישות הזיכרון במקרה של *קצב עדכון קבוע* או להגדיל בצורה לא מבוקרת את קצב התמסורת במקרה של *גודל בלוק קבוע*. כיוון שמיקום המאורע האחרון n_M הוא זה שקובע את גודל החפיפה עם הבלוק הבא, יש לאלץ אותו להמצא קרוב לסוף בלוק האנליזה.

סוגיה נוספת בבחירת הסגמנטציה ההתחלתית היא בחירת קצות הבלוק האפקטיבי שעבורו תמוזער השגיאה. באופן טבעי, תחילת הבלוק היא תמיד הוקטור הראשון (שהוא גם המאורע האחרון של הבלוק הקודם, עקב החפיפה). את הווקטור האחרון ב-ORTD בוחרים "מחוץ לבלוק" והוא למעשה וקטור אפס. פתרון כזה לא מתאים לחיפוש תת-אופטימלי, כי וקטור האפס ימשוך לעברו את המאורע האחרון כבר מהאיטרציה השנייה וזה ימוקם בקצה של הבלוק, דבר שיגרום לשגיאות מודל גדולות וביצועים נחותים של האלגוריתם התת-אופטימלי (ר' פרטים על האלגוריתם בהמשך סעיף זה).

אנו נבחן בהמשך שתי אופציות למיקום הקצה האפקטיבי : לבחור אותו בקצה האמיתי של הבלוק או לבחור אותו לקבלת מינימום גלובלי של SFTR עבור $L \approx \tilde{N}/M$ מסגרות אחרונות של בלוק האנליזה (נתייחס לשיטה זו כאל *SFTR מוגבל* (*L-confined SFTR*). ניתן במקרה השני להגדיר את הקצה האפקטיבי של בלוק האנליזה כמאורע האחרון, כלומר לא לשנות את המיקום ההתחלתי המאורע האחרון בשלב של החיפוש. אפשרויות שונות של הסגמנטציה ההתחלתית, שנבדקו מסוכמות בטבלה 5-א.

Table 5-a. Initial segmentations, examined for the SORTeD algorithm.

טבלה 5-א. אפשרויות שנבחנו לקביעת הסגמנטציה ההתחלתית עבור אלגוריתם ה-SORTeD

Initial Cond.	n_0	$n_0 - n_{M-1}$	n_M	n_{M+1}
1.	$n_0 = 1$	Uniform on $(n_0, n_{M+1}]$		By *L-confined SFTR*
2.	$n_0 = 1$	Uniform on $(n_0, n_{M+1}]$		$n_{M+1} = \tilde{N}$
3.	$n_0 = 1$	Uniform on (n_0, n_M)	By *L-confined SFTR*	$n_{M+1} = n_M$
4.	$n_0 = 1$	SFTR's local minima on (n_0, n_M)	By *L-confined SFTR*	$n_{M+1} = n_M$
5.	$n_0 = 1$	SFTR's local minima on (n_0, n_M)	By *L-confined SFTR*	$n_{M+1} = \tilde{N}$

אם הגבול האפקטיבי של הבלוק לא מתלכד אם המאורע האחרון (תנאי התחלה 1,2,4 בטבלה 3), אזי המיקום של המאורע האחרון יכול להשתנות ויש להבטיח בעת החיפוש, שמאורע זה לא יתרחק יותר מדי מהקצה הימני של הבלוק, דבר שעלול להביא להגדלת החפיפה הלא רצויה. אבל תנאי התחלה אלו צפויים להיות טובים מאחרים, כיוון שבהם גם המאורע האחרון יכול לשנות את מיקומו וכך להביא להקטנת השגיאה הכוללת לבלוק.

תוצאות הניסויים המעשיים (ר' סעיף 6.2.3) מעידים על כך, שהתצורות 1,2,5 בהן לא מקבעים את המאורע האחרון משיגות תוצאות טובות יותר משמעותית משאר התצורות (שיפור של 0.4-0.5 dB

של ה-LSD). הדבר מעיד על כך שאין לקבע את המאורעות, ולו אף את המאורע האחרון, לקבלת ביצועים מיטביים. תצורה מס' 2 הניבה תוצאות הטובות ביותר, ולכן היא נבחרה כברירת מחדל למערכת. מתוצאה זו ניתן להסיק כי האלגוריתם התת-אופטימלי אינו רגיש במיוחד לסגמנטציה ההתחלתית, פרט למיקום של קצה הבלוק. תנאי ההתחלה המבוססים על הקריטריון ליציבות ספקטרלית (תצורה מס' 5) לא היו טובים מהפיזור האחיד (תצורה מס' 2).

5.4.5 אלגוריתם תת-אופטימלי לשיפור הסגמנטציה ההתחלתית

האלגוריתם התת-אופטימלי בא להקטין את סיבוכיות החיפוש המלא (מסיבוכיות ריבועית לסיבוכיות לינארית ביחס לגודל הבלוק). אלגוריתם זה מסתמך על הסגמנטציה ההתחלתית ומבצע עידון שלה ע"י מינימיזציה מקומית של השגיאה בתת-בלוקים המכילים בדיוק 2 סגמנטים עוקבים.

תהי $\mathbf{Y}_{n_k}$ תת-מטריצה של Y אשר מכילה את העמודותיה מה- n_k עד ל- n_{k+1} $(1 \le k \le M)$.

בהנחה, שידועים וקטורי המאורע, נוכל לחשב את פונקציות המאורע (למשל, ע"י (5.3)) והשגיאה המצטברת לתת-מטריצה זו תהיה:

$$E_{block}\left(\mathbf{Y}_{n_k}\right) = \sum_{n=n_k}^{n_{k+1}-1} E(n) = E_{seg}(n_{k-1}, n_k) + E_{seg}(n_k, n_{k+1}) \qquad (5.12)$$

כאשר $E(n)$ נתון ע"י (5.11).

נניח שהגבולות של $\mathbf{Y}_{n_k}$ מקובעים. נוכל לבצע חיפוש של מרכז מאורע ה-k האופטימלי n_k^*, אשר מביא למינימום את השגיאה (5.12):

$$\begin{cases} E_{\min}(\mathbf{Y}_{n_k}) = \min\limits_{n_k \in (n_{k-1}, n_{k+1})} \left(E_{seg}\left(n_{k-1}, n_k\right) + E_{seg}\left(n_k, n_{k+1}\right)\right) \\ n_k^* = \arg\min\limits_{n_k \in (n_{k-1}, n_{k+1})} \left(E_{seg}\left(n_{k-1}, n_k\right) + E_{seg}\left(n_k, n_{k+1}\right)\right) \end{cases}, \ k = 1, \ldots M \qquad (5.13)$$

ע"י הפעלת החיפוש הנ"ל בצורה איטרטיבית, על כל מרכזי המאורעות בסדר עולה, נקבל את אלגוריתם החיפוש התת-אופטימלי. ניתן לעשות יותר ממעבר אחד לשיפור האלגוריתם (בד"כ שני מעברים מספיקים להתכנסות האלגוריתם). כפי שניתן לראות באיור 17 אלגוריתם החיפוש מתבצע מספר פעמים, כאשר בהפעלה ראשונה אין מסתמכים יתר על המידה על הסגמנטציה הראשונית, ולכן שינוי בסגמנטציה תוך כדי האלגוריתם גורר שינוי של וקטורי המטרה לפי (4.21). בהפעלות הבאות (לאחר עידון וקטורי המטרה), אין משנים יותר את וקטורי המטרה במהלך החיפוש. בטבלה 5-ב מפורט אלגוריתם זה לחיפוש הסגמנטציה עבור ההרצה הראשונה (וקטורי המטרה לא ידועים גם הם) ובטבלה 5-ג מסוכם אלגוריתם זה בהפעלות הבאות, כאשר אין משנים את וקטורי המטרה במהלך החיפוש.

Table 5-b. Sub-optimal algorithm for block segmentation search (inside SORTeD). The initial application (targets unknown)

טבלה 5-ב. האלגוריתם התת-אופטימלי למציאת הסגמנטציה לבלוק – הפעלה ראשונה (וקטורי המטרה לא ידועים)

אתחול	סגמנטציה ראשונית: $\left\{n_k^{curr}\right\}_{k=0}^{M+1} = \left\{n_k^{0}\right\}_{k=0}^{M+1}$
איטרציה כללית	$I \leftarrow 1$ FOR k = 1 to M-1: <u>צעד חיפוש:</u> $n_k^* = \underset{i\in(n_{k-1}^{curr},n_{k+1}^{curr})}{\arg\min}\ E_{block}(n_{k-1}^{curr}, i, n_{k+1}^{curr})$, כאשר $E_{block}(n_{k-1}^{curr}, i, n_{k+1}^{curr}) = \sum_{n=n_{k-1}}^{i-1} E(n, \mathbf{y}_{n_{k-1}^{curr}}, \mathbf{y}_i) + \sum_{n=i+1}^{n_{k+1}-1} E(n, \mathbf{y}_i, \mathbf{y}_{n_{k+1}^{curr}})$ ו- $E(n, \mathbf{a}_{left}^{(i)}, \mathbf{a}_{right}^{(i)}) = (\mathbf{y}_n - \mathbf{a}_{left}^{(i)}\phi_{left}^{(i)}(n) - \mathbf{a}_{right}^{(i)}\phi_{right}^{(i)}(n))^T \mathbf{W}_n (\mathbf{y}_n - \mathbf{a}_{left}^{(i)}\phi_{left}^{(i)}(n) - \mathbf{a}_{right}^{(i)}\phi_{right}^{(i)}(n))$ איפה ש- $\begin{pmatrix}\phi_{left}^{(i)}(n)\\ \phi_{right}^{(i)}(n)\end{pmatrix} = \begin{cases}\boldsymbol{\varphi}(\mathbf{y}_n, \mathbf{y}_{n_{k-1}^{curr}}, \mathbf{y}_i), & n_{k-1}^{curr} \le n < i\\ \boldsymbol{\varphi}(\mathbf{y}_n, \mathbf{y}_i, \mathbf{y}_{n_{k+1}^{curr}}), & i \le n < n_{k+1}^{curr}\end{cases}$ ו- $\boldsymbol{\varphi}$ הנו אופרטור חישוב אופטימלי לפי (5.3) או אופרטור אחר. $n_k^{curr} \leftarrow n_k^*$
עדכון	$n_k^{opt} \leftarrow n_k^{curr}, \quad 1 \le k \le M$ $\mathbf{a}_k^{opt} \leftarrow \mathbf{y}(n_k^{curr}), \quad 1 \le k \le M$ $\begin{pmatrix}\phi_{k-1}^{opt}(n)\\ \phi_k^{opt}(n)\end{pmatrix} \leftarrow \begin{pmatrix}\phi_{left}^{(n_k^{curr})}(n)\\ \phi_{right}^{(n_k^{curr})}(n)\end{pmatrix}, \quad n_{k-1}^{curr} \le n < n_k^{curr}, \quad 1 \le k \le M$ $\phi_M^{opt}(n) \leftarrow \phi_{left}^{(n_M^{curr})}(n), \quad n_M^{curr} \le n < n_{M+1}^{curr}.$
סיום	אם $I \le$ Number of iterations, $I \leftarrow I+1$, חזור לשלב האיטרציה אחרת, סיים.

Table 5-c. Sub-optimal algorithm for block segmentation search (inside SORTeD). Non-initial run (the targets are known).

טבלה 5-ג. האלגוריתם התת-אופטימלי למציאת הסגמנטציה של בלוק – הפעלה כלשהי, פרט לראשונית (וקטורי המטרה ידועים)

אתחול	סגמנטציה ראשונית: $\left\{n_k^{curr}\right\}_{k=0}^{M+1}=\left\{n_k^0\right\}_{k=0}^{M+1}$ וקטורי המטרה (לאחר העידון): $\left\{\mathbf{a}_k\right\}_{k=0}^{M+1}=\left\{\mathbf{a}_k^{(0)}\right\}_{k=0}^{M+1}$
איטרציה כללית	$I\leftarrow 1$ FOR k = 1 to M-1: <u>צעד חיפוש:</u> $n_k^{*}=\underset{i\in(n_{k-1}^{curr},n_{k+1}^{curr})}{\arg\min}\ E_{block}(n_{k-1}^{curr},i,n_{k+1}^{curr})$, כאשר $E_{block}(n_{k-1}^{curr},i,n_{k+1}^{curr})=\sum_{n=n_{k-1}}^{i-1}E(n,\mathbf{a}_{k-1},\mathbf{a}_k)+\sum_{n=i+1}^{n_{k+1}-1}E(n,\mathbf{a}_k,\mathbf{a}_{k+1})$ ו- $E(n,\mathbf{a}_{left}^{(i)},\mathbf{a}_{right}^{(i)})=(\mathbf{y}_n-\mathbf{a}_{left}^{(i)}\phi_{left}^{(i)}(n)-\mathbf{a}_{right}^{(i)}\phi_{right}^{(i)}(n))^T\mathbf{W}_n(\mathbf{y}_n-\mathbf{a}_{left}^{(i)}\phi_{left}^{(i)}(n)-\mathbf{a}_{right}^{(i)}\phi_{right}^{(i)}(n))$ איפה ש- $\begin{pmatrix}\phi_{left}^{(i)}(n)\\ \phi_{right}^{(i)}(n)\end{pmatrix}=\begin{cases}\boldsymbol{\varphi}(\mathbf{y}_n,\mathbf{a}_{k-1},\mathbf{a}_k), & n_{k-1}^{curr}\le n<i\\ \boldsymbol{\varphi}(\mathbf{y}_n,\mathbf{a}_k,\mathbf{a}_{k+1}), & i\le n<n_{k+1}^{curr}\end{cases}$ ו- $\boldsymbol{\varphi}$ הנו אופרטור חישוב אופטימלי לפי (5.3) או אופרטור אחר. $n_k^{curr}\leftarrow n_k^{*}$
עדכון	$n_k^{opt}\leftarrow n_k^{curr}$, $1\le k\le M$ $\begin{pmatrix}\phi_{k-1}^{opt}(n)\\ \phi_k^{opt}(n)\end{pmatrix}\leftarrow\begin{pmatrix}\phi_{left}^{(n_k^{curr})}(n)\\ \phi_{right}^{(n_k^{curr})}(n)\end{pmatrix}$, $n_{k-1}^{curr}\le n<n_k^{curr}$, $1\le k\le M$ $\phi_M^{opt}(n)\leftarrow\phi_{left}^{(n_M^{curr})}(n)$, $n_M^{curr}\le n<n_{M+1}^{curr}$.
סיום	אם $I\le$ Number of iterations, $I\leftarrow I+1$, חזור לשלב האיטרציה, אחרת, סיים.

5.4.6 השוואת הסיבוכיות של האלגוריתמים ORTD ו-SORTeD

נעריך כעת את הסיבוכיות של האלגוריתם התת-אופטימלי (SORTeD) ונשווה אותה ל-ORTD. ניזכר תחילה (ר' סעיף 5.3.2), כי בשלב של עידון וקטורי המטרה (שהוא זהה עבור ה-ORTD וה-SORTeD) מבצעים $pN/2$ פעולות, כיוון שיש לבצע $N/2$ צעדים בחילוצי גאוס (Gauss) ב-p מערכות משוואות תלת-אלכסוניות סימטריות, כאשר N הוא אורך הבלוק ו-p הוא מספר הרכיבים בוקטורי הפרמטרים.

נעריך כעת את מספר הפעולות הקריטיות בשלב של חיפוש הסגמנטציה עבור שני האלגוריתמים, כגון מספר ההשוואות ומספר החישובים של השגיאה הרגעית. חישובים אלו נחוצים לקביעת שגיאות הסגמנטים והם כרוכים בחישוב של פונקציות המאורע הרגעיות המתאימות.

בכל הפעלה של האלגוריתם התת-אופטימלי, ישנו סדר גודל של $2N$ השוואות (כיוון שיש לבחון M המאורעות, ובכל אחת מהבדיקות נבדקים $2N/M$ מקומות בממוצע).

לעומת זאת, בחיפוש המלא של הסגמנטציה מספר ההשוואות הוא כמספר הקשתות בדיאגרמת ה-trellis (ר' איור 4-י), כלומר ניתן להעריכו ע"י $MN^2/2$.

נעריך כעת כמה פעמים יש לחשב שגיאה רגעית (שגיאת סגמנט היא סכום של השגיאות הרגעיות של המסגרות שהסגמנט מכיל). נבחין בין שני מקרים: ההפעלה הראשונית של שלב חיפוש הסגמנטציה וההפעלות הבאות של שלב זה. ההבדל העיקרי ביניהם הוא בכך, שבהפעלה ראשונית לא יודעים מה הערך של וקטורי המטרה ומשתמשים בהנחה של $\mathbf{a}_k = \mathbf{y}(n_k)$, כאשר $\mathbf{a}_k$ הוא וקטור המטרה ה-*k*-י ו-n_k הוא מרכז המאורע ה-*k*-י. לכן כל שינוי בסגמנטציה במהלך החיפוש מצריך חישוב מחודש של השגיאות הרגעיות בסגמנטים הסמוכים למקום השינוי. בהפעלות הבאות לא משנים את ערך וקטורי המטרה במהלך שלב החיפוש, לכן יש לבצע פחות חישובים אלה.

בהפעלה ראשונית של שלב חיפוש הסגמנטציה, כאמור, יש צורך בכל צעד חיפוש (ויש $2N$ כאלה) לחשב מחדש את השגיאות הרגעיות הרלבנטיות. באלגוריתם התת-אופטימלי יש לעשות סדר גודל של $2N/M$ חישובים כאלה בממוצע (כמספר המסגרות הממוצע בשני הסגמנטים הסמוכים). כלומר בהפעלה ראשונית של האלגוריתם מבצעים כ- $\left(2N/M\right)2N = 4N^2/M$ חישובים של השגיאה הרגעית. לעומת זאת, בהפעלות הבאות, כאשר וקטורי המאורע לא משתנים בצעדי החיפוש, מספיק לחשב פעם אחת את כל השגיאות הרגעיות, ואז בכל שינוי מיקום המאורע רק 2 מסגרות הקצה ישנו את ערך השגיאה הרגעית שלהן, וניתן להגיע לשגיאה מצטברת חדשה ע"י החסרת שגיאות קודמות והוספת שגיאות חדשות. כלומר, מבצעים כ- $N + \left(2 \times 2N/M\right)M = 5N$ חישובים של שגיאה רגעית.

נאמוד כעת מספר החישובים של השגיאה הרגעית ב-ORTD. בהפעלה ראשונה של חיפוש הסגמנטציה המלא, כל מסגרת (מתוך N) יכולה להיכלל בכ- $N^2/2$ סגמנטים שונים, ולכן יש לחשב כ-

$N^3/2$ שגיאות רגעיות. לעומת זאת, בהפעלות הבאות (לאחר עידון וקטורי המטרה) יש לחשב רק כ-M שגיאות רגעיות לכל מסגרת, כיוון שוקטורי המטרה אינם משתנים, והשגיאה הרגעית תשתנה רק אם ישתנה שיוך של מסגרת מסוימת (כשהיא עוברת מהסגמנט ה-i לסגמנט ה-$i+1$). כלומר, ישנם כ-MN חישובים של השגיאה הרגעית.

נשים לב, כי הסיבוכיות של חישוב השגיאה הרגעית היא $O(p)$ ומספר הפעולות המדויק תלוי באיזה מהפתרונות לפונקצית המאורע הרגעית משתמשים לחישוב השגיאה (ר׳ סעיף 5.4.7).

נסכם את החסמים הנ״ל על מספר החישובים בטבלה 5-ד. (שים לב כי נתון מספר פעולות לבלוק בעל N מסגרות לערך, כלומר יש לחלק את המספרים ב-N לקבלת סיבוכיות הפעולות למסגרת בודדת.) חישוב של קירובים טובים יותר של מספר ההשוואות ומספר החישובים של השגיאה הרגעית אפשר למצוא בנספח II. בטבלה זו מוצגים גם ערכים מדויקים אלה עבור M=3 ו-N=11 (התצורה שנבחרה במקודד הדיבור הסופי).

Table 5-d. Suboptimal (SORTeD) vs. optimal (ORTD) segmentation complexity (per *N*-frame block with *M* events)

טבלה 5-ד. סיבוכיות של האלגוריתם האופטימלי והתת-אופטימלי עבור בלוק המכיל N מסגרות ו-M מאורעות.

SORTeD	ORTD		הפעולות	
$2N$	$N^2M/2$	החסם	מס׳ השוואות	
12	72	M=3, N=11		
$4N^2/M$	$N^3/2$	החסם	הפעלה ראשונית	מס׳ חישובי שגיאה רגעית
27	264	M=3, N=11		
$5N$	MN	החסם	הפעלות אחרות	
17	33	M=3, N=11		

5.4.7 דרכים לחישוב פונקציות מאורע רגעיות

אם משתמשים בפירוק ה-TD לקידוד בקצבים הנמוכים ביותר, אזי חשוב שפונקציות המאורע יהיו בעלות צורה פשוטה הניתנת לקידוד VQ במספר מצומצם של סיביות ללא פגיעה משמעותית בביצועים. ראינו כבר (איור 4-ו, איור 4-ט) כי צורות של פונקציות המאורע של RTD (גם ללא אילוץ, גם המשלימות לאחד) אינן רגולריות, לעיתים, ולכן קשות לקוונטיזציה ווקטורית. לצורך פישוט, ניתן להמיר הפתרון האופטימלי ב-(5.3) בפתרונות מאולצים, דוגמת (4.23) והצעות נוספות שיידונו בהמשך. ברור, כי הכנסת אילוצים כלשהם תגדיל את השגיאה כוללת, אך שילובם באלגוריתמי חיפוש עשוי להקטין את הדגרדציה.

מספר אילוצים הוצעו בספרות [40,42]. אנחנו נכליל ונשפר את האילוצים המוצעים וכן נציג את הביטויים עבור פתרונות עם קריטריון של WSE. האילוצים שמוכנסים לתוך חישוב פונקצית המאורע הן:

1) אילוץ ההשלמה לאחד של זוג הפונקציות הרגעיות
2) אילוץ האי-שליליות
3) אילוץ של ערך גבוה של פונקצית המאורע במרכז המאורע (מרכוז)
4) אילוץ מונוטוניות

בהמשך הסעיף נפרט לגבי כ״א מהאילוצים הללו.

האילוץ של השלמה לאחד (4.22), אשר הוצג בשילוב עם אילוצים נוספים ב-[42,40], מצמצם פי שניים את כמות המידע שיש לקודד. הכללה מידית של הפתרון ב-(4.23), עבור שגיאה ריבועית משוקללת (WSE), תינתן ע״י:

$$\begin{pmatrix} \bar{\phi}_k(n) \\ \bar{\phi}_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \dfrac{\left(\mathbf{y}(n)-\mathbf{a}_{k+1}\right)^T \mathbf{W}(n)\left(\mathbf{a}_k-\mathbf{a}_{k+1}\right)}{\left(\mathbf{a}_k-\mathbf{a}_{k+1}\right)^T \mathbf{W}(n)\left(\mathbf{a}_k-\mathbf{a}_{k+1}\right)} \\ 1-\bar{\phi}_k(n) \end{pmatrix}, \; n_k \le n < n_{k+1} \tag{5.14}$$

Kim *et al* מחמירים את האילוץ (4.22) בצורה הבאה [42]:

$$\begin{cases} \phi_k(n)+\phi_{k-1}(n)=1, \\ 0 \le \phi_k(n) \le 1, \\ \phi_k(n_k)=1. \end{cases} \tag{5.15}$$

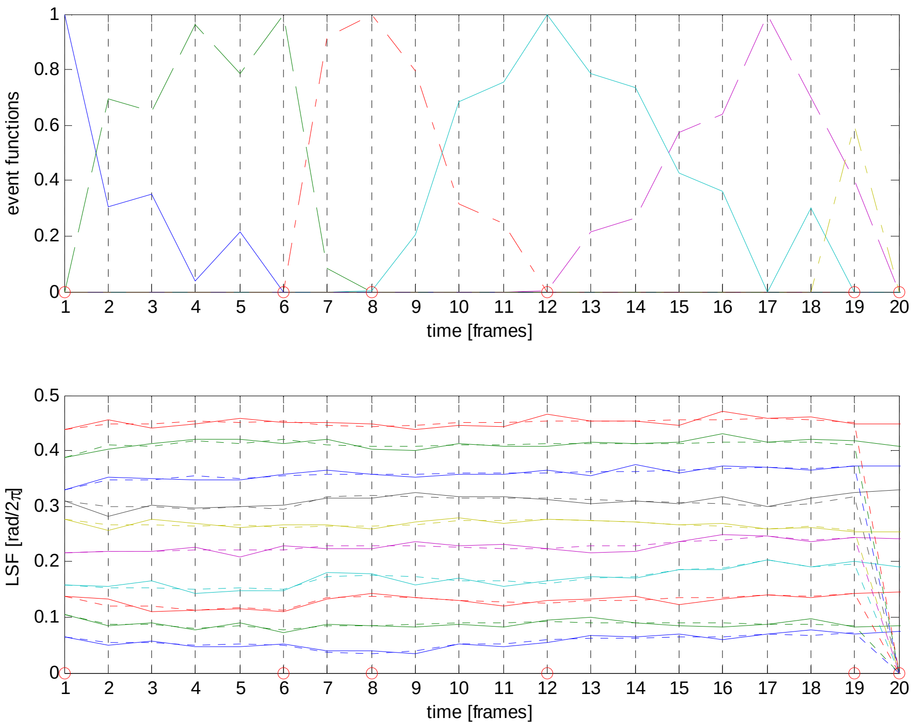
מבחינה גיאומטרית שני האילוצים הראשונים ב-(5.15) מביאים לקירוב וקטורי הפרמטרים ע״י ההטלה לתוך *קטע* של הישר המחבר בין וקטורי המטרה שמצידיהם. כמו כן, האילוצים האלה

מבטיחים שמירה על תכונת הסדר של מקדמי ה-LSF (ר' פרק 2.3.3) וכן מגבילים את התחום הדינמי של פונקציות המאורע. האילוץ השלישי $(\phi_k(n_k)=1)$ מהווה למעשה דרישת המרכוז של פונקצית המאורע סביב מרכזה, בו "ממוקם" וקטור המטרה. מזעור של (4.19) בהינתן וקטורי המאורע והאילוצים ב-(5.15) נותן [40]:

$$\hat{\phi}_k(n)=\begin{pmatrix} 1, & n=n_k \\ \min(1,\max(0,\overline{\phi}_k(n))) & n_k\le n<n_{k+1} \\ 1-\hat{\phi}_{k-1}(n), & n_{k-1}\le n<n_k \\ 0, & else \end{pmatrix}, \qquad (5.16)$$

כאשר $\overline{\phi}_k(n)$ חושב ע"י (5.14). הפתרון ניתן בצורה רקורסיבית, אך בפועל, מספיק לחשב את הענפים הימניים של פונקציות המאורע $(\hat{\phi}_k(n), n_k\le n<n_{k+1})$, והענפים השמאליים $(\hat{\phi}_k(n), n_{k-1}\le n<n_k)$, החופפים איתם פשוט ישלימו אותם לאחד בכל רגע *n*.

פונקציות המאורע המתוארות ב-(5.16) מוצגות באיור 5-ד. העובדה שהענפים של פונקציות המאורע אינם מונוטוניים, עלולה להקשות על הקוונטיזציה של צורות אילו עקב קיום מספר שיאים מקומיים. (ניתן להבחין במספר "אונות צד" לפונקציות מאורע אחדות, בדומה לאלו המוצגים באיור 4-ה) .



איור 5-ד. פירוק RTD עם פונקציות מאורע המשלימות לאחד, אי שליליות וממורכזות.
(1) פונקציות מאורעות מאולצות עבור פירוק RTD עם סגמנטציה נתונה (קרי, מרכזי מאורעות ידועים). מרכזי המאורעות מסומנים בעיגולים. (2) קירוב ל-LSF, שהתקבל ע"י פירוק ה-RTD המאויר ב-(1). ה-LSF המקוריים מסומנים בקווים רציפים, וה-LSF הממודלים מסומנים בקווים מקווקווים. מרכזי המאורעות מסומנים בעיגולים.

Figure 5-4. RTD with one's complementary non-negative and centered event functions. (1) The constrained event functions for RTD with given segmentation (i.e. event centers). Event centers are shown by circles. (2). LSF approximation, obtained by RTD decomposition, plotted in (1). Original LSF trajectories are plotted by solid lines, while modeled ones are plotted by dotted lines. Event centers are shown by circles.

אילוץ המרכוז ($\phi_k(n_k)=1$) עלול להיות מחמיר מדי, שכן וקטורי המטרה לאחר העידון שונים מווקטור הפרמטרים, אשר נמצא במרכז המאורע. לצורך החלשה של האילוץ הזה, נציג פתרון (5.16) כהרכבה של שלושה אילוצים – אילוץ השלמה לאחד, אילוץ החיוביות ולבסוף אילוץ המרכוז. בלי הגבלת הכלליות, נתייחס מכאן ואילך לפתרון עבור הענפים הימניים בלבד של פונקציות המאורע, מכיוון שהענפים השמאליים משלימים אותם לאחד בכל רגע ורגע.

אילוץ ההשלמה לאחד מביא לפתרון הניתן ב-(5.14). הוספת אילוץ החיוביות מביאה ל:

$$\phi^{*}_{k}(n)=\min(1,\max(0,\bar{\phi}_{k}(n))),\quad n_k \le n < n_{k+1}, \tag{5.17}$$

כאשר $\bar{\phi}_k(n)$ חושב ע"י (5.14).

לאילוץ החיוביות נוסיף עתה את אילוץ המרכוז המוחלש. נדרוש, כי ערך מרכזי יהיה בטווח $[1-\tau,1]$:

$$\phi^{**}_{k}(n)=\begin{cases}\phi^{*}_{k}(n), & 1-\tau \le \phi^{*}_{k}(n)\le 1\\ 1-\tau, & \phi^{*}_{k}(n)<1-\tau\end{cases} \tag{5.18}$$

ב-[42] מציעים פתרון, אשר מבטיח, בנוסף לאילוצים (5.15), מונוטוניות של כל אחד משני הענפים של פונקציות המאורע המתקבלות:

$$\tilde{\phi}_k(n)=\begin{pmatrix} 1, & n=n_k\\ \min(\hat{\phi}_k(n),\tilde{\phi}_k(n-1)) & n_k\le n<n_{k+1}\\ 1-\tilde{\phi}_{k-1}(n), & n_{k-1}\le n<n_k\\ 0, & else\end{pmatrix}, \tag{5.19}$$

כאשר $\hat{\phi}_k(n)$ חושב ע"י (5.16).

אילוץ זה מפשט עוד יותר את צורת פונקציות המאורע ע"מ להקל על קידוד וקטורי של צורות פונקציות המאורע (ר' איור 5-ה).

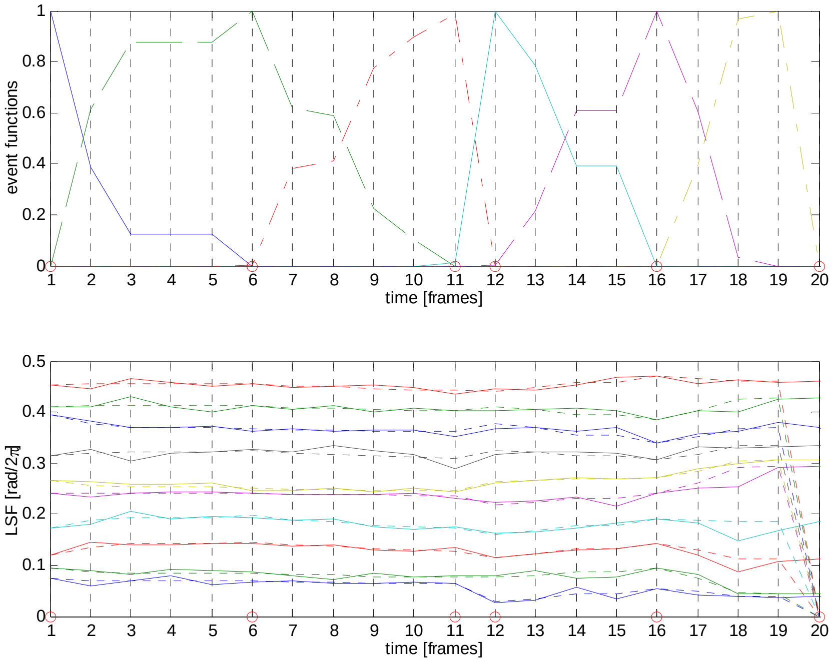


Figure 5-5. RTD with one's complementary, non-negative, centered and monotonic event functions branches. (1) The constrained event functions for RTD with given segmentation (i.e. event centers). Event centers are shown by circles. (2) LSF approximation, obtained by RTD decomposition, plotted in (1). Original LSF trajectories are plotted by solid lines, while modeled ones are plotted by dotted lines. Event centers are shown by circles.

איור 5-ה. פירוק RTD עם פונקציות מאורע המשלימות לאחד, אי שליליות, ממורכזות ומונוטוניות. (1) פונקציות מאורעות מאולצות עבור פירוק RTD עם סגמנטציה נתונה (קרי, מרכזי מאורעות ידועים). מרכזי המאורעות מסומנים בעיגולים. (1). קירוב ל-LSF, אשר התקבל ע״י פירוק ה-RTD המאויר ב-(2) ה-LSF המקוריים מסומנים בקווים רציפים, וה-LSF הממודלים מסומנים בקווים מקווקווים. מרכזי המאורעות מסומנים בעיגולים.

אילוץ זה יכול לעתים להביא להגדלת שגיאה משמעותית, במיוחד כאשר הענף הימני של פונקצית המאורע האופטימלית מכיל ירידה תלולה ואחר כך עליה חדה. במקרה כזה האילוץ יביא לערכים נמוכים לאחר הירידה ועד הסוף. אנו מציעים אלגוריתם משופר למציאת פונקציות מאורע מונוטוניות אשר מביא בהכרח לשיפור ב-WSE לעומת השיטה של [43]. האלגוריתם מופעל על הענף הימני (היורד) של פונקצית מאורע ומשהה את ההחלטה על הצורה הסופית שלה. האלגוריתם דורש חישובים נוספים של השגיאה הרגעית (5.11), אך החישובים האלה מתבצעים רק ב-12% מכלל פונקציות מאורעות [42]. האלגוריתם מופעל למעשה רק ברגעים שהפתרון האופטימלי מפר את דרישת המונוטוניות. במקרה שהפתרון האופטימלי מפר את תכונת המונוטוניות, שתי אפשרויות של פונקצית המאורע המונוטונית נבדקות, וזאת שמביאה לשגיאה המינימלית, נבחרת.

יהי $n_k < n < n_{k+1}$, $\boldsymbol{\varphi}_k(n) \triangleq [\phi_k(n_k), \phi_k(n_k+1), ..., \phi_k(n)]^T$ הענף הימני החלקי עד לרגע n (שאמור להיות מונוטוני יורד) של פונקצית המאורע ה-k. האלגוריתם המוצע ידאג כי $\boldsymbol{\varphi}_k(n)$ ישמור תמיד על תכונת המונוטוניות, כלומר $\phi_k(n_k) \geq \phi_k(n_k+1) \geq ... \geq \phi_k(n), \quad n_k < n < n_{k+1}$. אם ברגע n הפתרון הרצוי $\phi_k^{**}(n)$ מקיים את תכונת המונוטוניות, כלומר קטן או שווה לרכיב האחרון של $\boldsymbol{\varphi}_k(n-1)$ $\left(\phi_k^{**}(n) \leq \left[\boldsymbol{\varphi}_k(n-1)\right]_{n-1}\right)$, אזי הוא מוכנס לתוך ה-$\boldsymbol{\varphi}_k(n)$ $\left(\boldsymbol{\varphi}_k(n) = \left[\boldsymbol{\varphi}_k(n-1) \quad \phi_k^{**}(n)\right]\right)$, אחרת שני פתרונות אפשריים נבחנים :

$$\begin{bmatrix} \boldsymbol{\varphi}_k^{(1)}(n) = \left[\boldsymbol{\varphi}_k(n-1) \quad \left[\boldsymbol{\varphi}_k(n-1)\right]_{n-1}\right] \\ \boldsymbol{\varphi}_k^{(2)}(n) = \left[\max\left(\boldsymbol{\varphi}_k(n-1), \phi_k^{**}(n)\right) \quad \phi_k^{**}(n)\right] \end{bmatrix} \tag{5.20}$$

ונבחר הפתרון שמביא לשגיאת הסגמנט $E_{seg}(n_k, n)$ הקטנה מבין השניים. באיור 5-ו (הנגזר מסימולציה אמיתית) מודגמת הפעולה של האלגוריתם החדש בהשוואה לאילוץ מתוך [43]. האלגוריתם הישן בוחר תמיד את הירידה התלולה ביותר (בענף הימני של פונקצית המאורע) , לעומת האלגוריתם החדש, שבוחר את המסלול הממזער את שגיאת ה-WSE[1]. בדוגמה זו זהו המסלול העליון. בסימולציות, שימוש בתנאי המונוטוניות החדש שיפר את ה-LSD בכ-0.01 dB.

הצגנו, אפוא, שיטות יעילות לחישוב פונקציות מאורעות הממזערות את קריטריון השגיאה הרצוי ועם זאת בעלי צורה הנוחה לקוונטיזציה. ברור שהאילוצים גורעים מהפתרון האופטימלי, אך אנו מצפים שהביצועים של המערכת עם הפונקציות המאולצות לא ירדו בהרבה כתוצאה מקוונטיזציה של צורות המאורעות. בנוסף, שימוש בפונקציות מאורעות רגולריות מביא לסגמנטצית דיבור רובסטית, אשר יכולה לשמש ביישומי דיבור אחרים, בנוסף לקידוד, כמו זיהוי, סינטזה ועוד.

[1] המזעור באלגוריתם החדש הוא רק על פני המסלולים הנבדקים, אשר בהכרח יכילו את המסלול שנבחר באלגוריתם הישן.

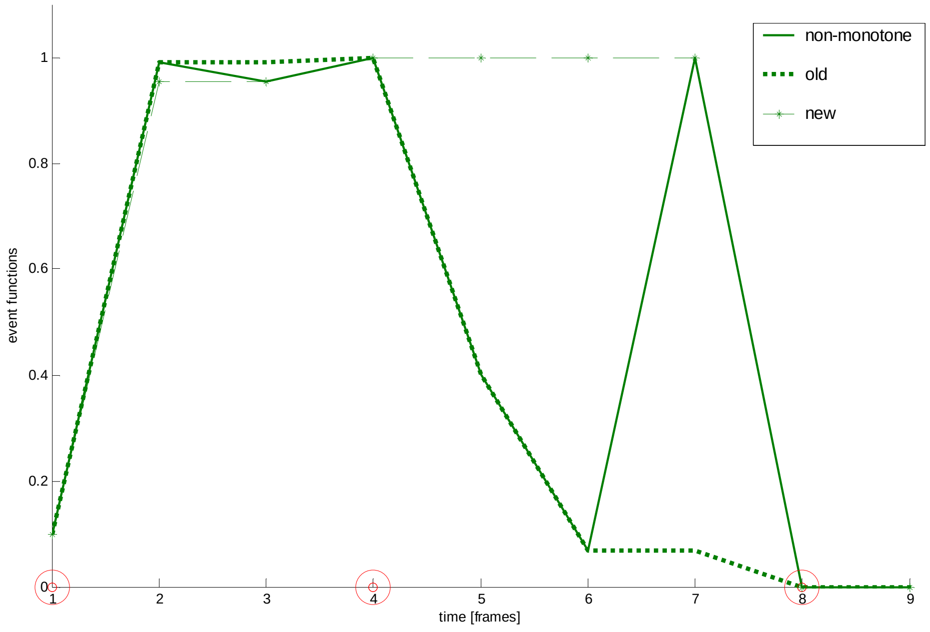


Figure 5-6. Improved monotony constraint. Ill-formed event function, fixed by monotone constraint imposing algorithms is plotted. Event centers are shown by circles. For the sake of simplicity, adjacent event functions (one's complement) are not plotted.

איור 5-6. אילוץ מונוטוניות משופר. דוגמה לתיקון של פונקצית המאורע בעלת "אונת צד" לא רצויה ע"י האלגוריתמים שמאלצים מונוטוניות. מרכזי המאורעות מסומנים בעיגולים. למען פשטות הציור, פונקציות המאורעות השכנות (המשלימות לאחד) לא משורטטות.

5.5 קוונטיזציה של פרמטרי ה-TD.

בסעיף זה נדון ביישום של אלגוריתם לפירוק ה-TD המוצע לשם דחיסת הפרמטרים הספקטרליים (LSF) בקצבים נמוכים ביותר.

הפרמטרים שיש לקוונט הם וקטורי המטרה ופונקציות האינטרפולציה. את וקטורי המטרה ניתן לקודד בדרכים המקובלות לפרמטרים ספקטרליים הממודלים, כיוון שעידון וקטורי המטרה לא משנה אותם משמעותית, וניתן להתייחס לוקטורי המטרה כאל וקטורי פרמטרים ספקטרליים נבחרים. השיטה שנבחרה בעבודה זו היא Split-VQ [13]. בשיטה זו מבצעים קוונטיזציה ווקטורית למספר תת-וקטורים בנפרד. בעבודה המקורית [13] נטען, כי הפירוק של וקטורי ה-LSF ל-4 וקטורים נמוכים ו-6 וקטורים גבוהים מביא לתוצאות משביעות רצון, כאשר ניתן תוך שימוש במשקלות PA-WSE לקבל ב-24 סיביות למסגרת איכות הקידוד שקופה.

עבור פונקציות האינטרפולציה, יש צורך לקודד בנפרד את אורכיהן ואת צורותיהן. עקב האילוצים אשר מופעלים על פונקציות המטרה, צורתם הכללית נוחה מאד לקידוד. ראשית, יש לקודד רק ענף

ימני (או לחילופין שמאלי) של פונקציות המאורע. בנוסף, ענפים אלה הם מונוטוניים ובעלי תחום דינמי דומה. כיוון שההערכה (המבוססת על הסימולציות) היא שהגודל האופייני של ספרי הקוד אלו היא 2-4 סיביות, ניתן ליצור ספרי קוד נפרדים לכל אורך מאורע אפשרי. לצורך קוונטיזציה נוחה של אורכי הסגמנטים, ההגבלה על אורך הסגמנט האפשרי הוכנסה כבר בשלב של חיפוש הסגמנטציה. אורך הסגמנט נע מ-1 עד 8. עקב כך, ישנם 8 ספרי קוד של צורות פונקציות המאורע. ספרי קוד אילו מתקבלים ע"י ביצוע של אלגוריתם ה-LBG על הענפים היורדים של פונקציות המאורע המונוטוניות, בנפרד לכל אורך אפשרי. כמובן שאפשר גם להשתמש בספר קוד אחיד ע"י שינוי ציר התדר של פונקציות המאורעות, אך הדבר עלול להגדיל את שגיאת קוונטיזציה והוא גם כבד מבחינה חישובית (עומס בזמן-אמת בקוונטיזציה, כי יש להמיר כל ווקטור בספר קוד משותף לאורך האמיתי בזמן החיפוש בספר הקוד לצורכי הקידוד).

ע"מ לקבל ביצועים מיטביים של אלגוריתם ה-TD, הן הקוונטיזציה של וקטורי המטרה והן הקוונטיזציה של פונקציות המאורע, משולבים בתהליך ה-TD. למעשה, הפתרון האופטימלי של פונקציות המאורע הרגעיות מומר במעבר על כל מילות קוד בספר הקוד המתאים, והצבתן לביטוי לשגיאת סגמנט.

ניתן להעריך את מספר הפעמים שיש לבצע חיפוש בספר קוד עבור הקידוד של בלוק באורך N. נראה כעת כי הוא פרופורציוני לאורך הבלוק N. ניזכר (סעיף 5.4.6), כי בכל צעד עדכון הסגמנטציה בודקים $2N/M$ מיקומים של מרכז המאורע. בכל בדיקה כזו יש למצוא את 2 הענפים האופטימליים (משני צידי מרכז המאורע, אשר עובר כוונון). סה"כ נעשים $(2\times {}^{2N}\!/_{M})M = 4N$ מעברים על ספרי קוד של פונקציות המאורע, שהיא פחות ממספר החישובים של פונקציות מאורע רגעיות, הן בהפעלה ראשונית והן בהפעלות הבאות של החיפוש התת-אופטימלי של הסגמנטציה, כחלק מאלגוריתם ה-SORTeD (כפי שחושב בסעיף 5.4.6).

כימות וקטורי המטרה משולב גם הוא בתהליך ה-TD. בהפעלה ראשונית של הסגמנטציה התת-אופטימלית, וקטורי היעד מקבלים ערכים של וקטורי הפרמטרים המקוונטים, אך השגיאה מחושבת ביחס לוקטורים המקוריים. בנוסף, לאחר עידון וקטורי המטרה הם עוברים את שלב הקוונטיזציה לפני שממשיכים בעדכון הסגמנטציה.

יש לציין, כי לא מובטח שיפור כתוצאה מעידון וקטורי המטרה והקוונטיזציה, לכן ערך ה-WSE מחושב לפני ואחרי העידון והקוונטיזציה, ואם אין שיפור, האלגוריתם מסתיים ומחזיר את האינדקסים לפונקציות המאורע ולוקטורי המטרה, אשר ממזערים את ה-WSE הכולל בבלוק.

5.6 סיכום

בפרק זה הוצע אלגוריתם לייצוג ודחיסה של רצף פרמטרים ספקטרליים של הדיבור. אלגוריתם זה נראה מתאים ליישומי זמן אמת (עם השהיות ארוכות יחסית). בחינת ביצועי האלגוריתם בנפרד ובשילוב עם מקודד MELP תקני יוצג בפרק הבא.

האלגוריתם התבסס על סכימת ה-ORTD (ר' סעיף 4.5) ומרחיב אותה, לשימוש בקריטריון שגיאת WMSE עם משקלות התלויות בווקטורי הכניסה. המודיפיקציה באה לידי ביטוי בשינויים בחישוב של פונקציות המאורע הרגעיות ועידון וקטורי המטרה. שילוב המשקלים לא העלה את סיבוכיות האלגוריתם.

כמו כן, על בסיס ה-ORTD פותח אלגוריתם תת-אופטימלי אשר מוריד בצורה משמעותית את הסיבוכיות של ה-ORTD (ר' טבלה 5-ד). טכניקה זו, הנקרית SORTeD, ניתנת ליישום במערכות זמן אמת.

ע"מ לאפשר קוונטיזציה עמוקה של פונקציות המאורעות, נעשה שימוש בפתרונות מאולצים לפונקציות המאורע הרגעיות לצורך בניית ספרי הקוד של צורות פונקציות המאורע. קוונטיזציה של וקטורי המטרה ופונקציות המאורעות שולבו בתוך אלגוריתם ה-SORTeD לשיפור התאמת המודל.

האלגוריתם שהוצע הוא כללי ויכול לשמש לדחיסה וייצוג של רצף וקטורי כלשהו. התמיכה בשגיאות ריבועיות המשוקללות עם משקלים התלויים בוקטורי הכניסה, מאפשרת הרחבת השימוש באלגוריתם זה גם לקידוד של הפרמטרי עירור אחדים (כפי שנראה בהמשך העבודה).

פרק 6

בחינת ביצועים של אלגוריתם ה-DW-SORTeD

6.1 מבוא

מודל ה-TD פותח במקור ככלי לאנליזה של אות דיבור, לכן עדכון וקטורי הפרמטרים היה צפוף מאד (כל 5 מילישניות) [36]. אנחנו, לעומת זאת, מתמקדים בעבודה הנוכחית בשימוש במודל ה-TD לדחיסת פרמטרים ספקטרליים עמוקה מאוד (עד 300 סיביות לשנייה). כעקרון, קצב עדכון וקטורי הפרמטרים הספקטרליים אינו משפיע ישירות על קצב פרמטרי ה-TD באלגוריתם DW-SORTeD, כיוון שהאלגוריתם שומר על קצב מאורעות לשנייה קבוע. אולם הגדלת קצב עדכון הפרמטרים הספקטרליים תגביר משמעותית את כמות החישובים שיש לבצע ביחידת זמן (אורך בלוק אנליזה, N, יגדל בהתאם) ולא ברור אם אילוצי הקצב הנמוך יאפשרו לאלגוריתם לתאר שינויים עדינים במסגרות עוקבות צפופות. אי לכך, נבחר לבצע את פירוק ה-TD על וקטורי הפרמטרים שמתעדכנים בהתאם לדרישות מקודדי הדיבור עליהם נתבסס.

מודל ה-ORTD נוסה בהצלחה בשילוב עם מקודד MELP-2400 התקני [6]. קצב העדכון הנמוך של פרמטרי ה-LSF בתקן זה (44.444 Hz) מאפשר להגיע לקצבים נמוכים מאוד בקידוד המעטפת הספקטרלית וזמינותו תאפשר בדיקת ביצועים של מערכת קידוד המעטפת הספקטרלית (באמצעות ה-TD) בשילוב עם קידוד העירור התקני.

הבעיה העיקרית בשילוב ה-TD במקודדי דיבור מעשיים היא ההשהיה הגבוהה שנדרשת ע״מ לקבל ביצועים סבירים. אורך בלוק האנליזה N, הוא זה שקובע בצורה הישירה את השהית המערכת. בהמשך הפרק נדגים, כי ניתן להשתמש באלגוריתם ה-DW-SORTD עם השהיות אלגוריתמיות של החל מ-7 מסגרות בלבד (כ-160 מילישניות), למרות שלא ניתן לנצל את מלוא העוצמה של המודל בהשהיות אילו. בהשהיה אלגוריתמית של 11 מסגרות (כ-250 מילישניות) ניתן להגיע לביצועים משופרים.

וקטורי הפרמטרים הספקטרליים המוכנסים ל-מודול ה-TD הם וקטורי ה-LSF במימד $p = 10$, כמקובל במקודדי דיבור פרמטריים לקצב נמוך, המאפיינים מסנני חיזוי לינארי, ששוערכו בשיטת האוטוקורלציה (ר׳ סעיף 2.3.2).

אותות הדיבור ששימשו בניסויים, נלקחו מבסיס הנתונים TIMIT (קצב דגימה מקורי של 16KHz), סוננו ע״י מסנן מעבר נמוכים עם תדר קיטעון של- 4KHz ולאחר מכן בוצעה המרת קצב ל-8KHz. אנליזת ה-LPC למציאת פונקצית האוטוקורלציה נעשתה בעזרת חלון Hamming במסגרות של 22.5ms, כפי שנעשה במקודד העירור המעורב (MELP) התקני [6].

בפרק הנוכחי נבחן את הביצועים של האלגוריתם על פני 20 משפטים מתוך בסיס הנתונים TIMIT (מחציתם נאמר ע״י גברים ומחציתם ע״י נשים). מדד הביצועים הן למודל ה-TD הלא מקוונט, הן למודל ה-TD בשילוב עם קוונטיזציה נעשה באמצעות ה-LSD עם תחום אינטגרציה שנע בין 100 ל-3400 Hz (ר׳ סעיף 2.4.1). בנוסף, תוכנת ה-PESQ התקנית [59] שימשה להערכת ציון ה-MOS (Mean Opinion Score) למערכת משלבת קידוד המעטפת הספקטרלית באמצעות DW-SORTeD וקידוד אות העירור בהתאם לתקן MELP-2400 [6]. המשקלים שנבחנו לביצוע ה-TD באמצעות DW-SORTeD הם משקלי Paliwal-Atal משקלי Gardner ומשקלי Gardner שעברו הנחתה קבועה של התדרים הגבוהים (ר׳ סעיף 5.2.2).

בתחילת הפרק נעריך את הביצועים של המודל ללא קוונטיזציה עבור קצבים שונים של מאורעות לשנייה ונבחר את פרמטרי האלגוריתם המועדפים עבור המערכת עם הקוונטיזציה. לאחר מכן, נעריך את הביצועים של המערכת האמיתית, אשר משלבת קוונטיזציה בתוך ה-TD.

6.2 הערכת הביצועים של מודל ה-TD

6.2.1 הקדמה

בסעיף זה נעריך את הביצועים של DW-ORTD לעומת ORTD ונראה את חשיבות של השימוש בשגיאות משוקללות, המבטאות בצורה טובה יותר את מידת הקרבה הנתפסת ע״י האוזן האנושית. בנוסף, נראה את ההשפעה של החיפוש התת-אופטימלי, לעומת החיפוש האופטימלי (SORTeD לעומת ORTD).

לצורך ביצוע ההשוואה, אורך הבלוק נבחר להיות 15 מסגרות, אם לא נאמר אחרת. הדבר שקול לחוצץ באורך של כ-340 מילישניות. מספר האיטרציות של אלגוריתמי ה-TD נבחר להיות שווה ל-3.

6.2.2 ORTD לעומת DW-ORTD עם משקלות שונים

תוצאות ההרצה של ORTD ו-DW-ORTD עם משקלות שונים (דינמיים) מוצגות בטבלה 6-א ובגרף שבאיור 6-א ההרצות נעשו עבור בחירה של מספר מאורעות שונים: M=3,4,5,6,7,8,9, בתוך בלוק באורך של N=15. ניתן להסיק מהתוצאות, כי השימוש במשקלים דינמיים משפר בצורה ניכרת את ההתאמה של הספקטרום, בהשוואה לקריטריון ה-MMSE של ה-ORTD המקורי. השיפור המרבי של 0.3-0.4 dB (LSD) הושג בשימוש במשקלי Gardner המונחתים (G-WSE(2)). ניתן

לראות בנוסף, כי הביצועים של המשקלים הדינמיים השונים עקביים עם הגרף באיור 5-א, שמציג את תכונות ההתאמה של השגיאות הריבועיות עם משקלים שונים למדד ה-LSD.

נשים לב, כי ההבדלים בין השקלולים השונים נשחקים ככל שקצב המאורעות עולה והספקטרום מתאים לאיכות שקופה (ה-LSD הממוצע מתחת ל-1dB). נקודות העבודה האופייניות עבור מערכות הקידוד המעשיות נמצאות בין 12 ל-20 מאורעות לשנייה, בהתאם לשיטת הדחיסה של הפרמטרים הספקטרליים ודרישות האיכות של המערכת הסופית.

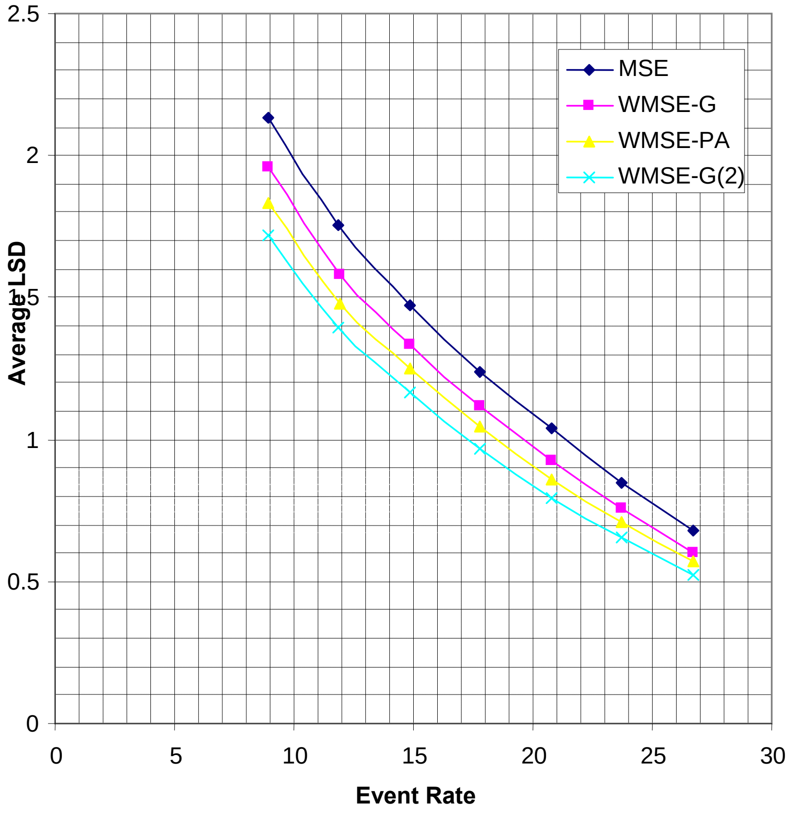


Figure 6-1. Spectral distortions (average LSD), obtained for different weighting of DW-ORTD's error criterion

איור 6-א. עיוותים ספקטרליים (ה-LSD הממוצע), אשר מתקבלים עבור שקלולים שונים של מדד השגיאה באלגוריתם ה-DW-ORTD

Table 6-a. Spectral distortion (average LSD and outliers percentage), obtained for different weighting of DW-ORTD error criterion. The numbers are shown for selected event rates and error weightings.

טבלה 6-א. עיוותים ספקטרליים (ה-LSD הממוצע ואחוזי החריגות), אשר מתקבלים עבור שקלולים שונים של מדד השגיאה באלגוריתם ה-DW-ORTD. מוצגות התוצאות המספריות של חלק קצבי המאורעות והשקלולים שנבחנו.

Algorithm	Ev./ sec	LSD		
		Avg., [dB]	2-4dB,[%]	>4dB,[%]
ORTD (no weighting)	11.85	1.7533	29.7598	4.5895
DW-ORTD (Paliwal-Atal)		1.4757	22.4292	1.7198
DW-ORTD (mod. Gardner)		1.3933	19.9928	1.3615
ORTD (no weighting)	14.81	1.4699	21.856	2.1856
DW-ORTD (Paliwal-Atal)		1.2514	15.9083	0.7166
DW-ORTD (mod. Gardner		1.1688	12.7911	0.6449
ORTD (no weighting)	17.78	1.24	16.1707	0.9322
DW-ORTD (Paliwal-Atal)		1.0485	9.896	0.4661
DW-ORTD (mod.Gardner)		0.9692	7.2427	0.2868

6.2.3 בחינת פרמטרים ל-DW-SORTeD

בטבלה 5-א בסעיף 5.4.4 הוצעו מספר תצורות אתחול לאלגוריתם התת-אופטימלי. הרצנו את האלגוריתם עם כל אחת מהתצורות הנ"ל ע"מ למצוא את התצורה בעלת הביצועים המיטביים. התוצאות ניתנות בטבלה 6-ב. תצורה מס' 2 (פיזור אחיד של כל המאורעות על פני הבלוק, מיקום של המאורע ה-$N+1$ בקצה של הבלוק) הניבה תוצאות מיטביות. ר' דיון על תוצאות אלו בסעיף 5.4.4.

Table 6-b. Experimental results for examining different initial segmentation setups of DW-SORTeD. The setup description may be found in Table 5-a.

טבלה 6-ב. תוצאות הניסויים לבחינת התצורות של תנאי ההתחלה של ה-DW-SORTeD. (ר' טבלה 5-א.)

Initial Setup No.	LSD		
	Avg., [dB]	2-4dB,[%]	>4dB,[%]
1.	1.5136	22.9473	2.5816
2.	1.4677	21.3185	2.1498
3.	1.7841	28.3769	4.4787
4.	1.8163	29.9534	4.6220
5.	1.4935	21.9433	2.4023

6.2.4 DW-ORTD לעומת DW-SORTeD

תוצאות ההרצה של ה-DW-ORTD עם משקלי Gardner מונחתים (שהניבו התאמה מיטבית עד כה) הושוו עם תוצאות ההרצה של האלגוריתם התת-אופטימלי DW-SORTeD, שהורץ עם אותם המשקלות (איור 6-ב, טבלה 6-ג). הדגרדציה שנגרמה עקב מעבר מחיפוש סגמנטציה מלא לחיפוש חלקי הייתה 0.05-0.07 dB בלבד, בטווח פעולה של 12-18 מאורעות לשנייה. הדגרדציה היחסית עולה כאשר מעלים את קצב המאורעות, אך אין לזה משמעות כי האיכות בקצבים אילו נשארת שקופה.

Table 6-c. Spectral distortion (average LSD and outliers percentage), obtained for DW-ORTD and DW-SORTeD. Modified Gardner weights were used for both.

טבלה 6-ג. עיוותים ספקטרליים (ה-LSD הממוצע ואחוזי החריגות), אשר מתקבלים עבור אלגוריתמי ה-DW-ORTD ו-DW-SORTeD. משקלי Gardner מונחתים שימשו בשני האלגוריתמים.

Algorithm	Ev./sec	LSD		
		Avg., [dB]	2-4dB,[%]	>4dB,[%]
DW-ORTD	11.85	1.3933	19.9928	1.3615
DW-SORTeD		1.4677	21.3185	2.1498
DW-ORTD	14.81	1.1688	12.7911	0.6449
DW-SORTeD		1.2100	13.0777	1.1107
DW-ORTD	17.78	0.9692	7.2427	0.2868
DW-SORTeD		1.0308	8.6411	0.7530

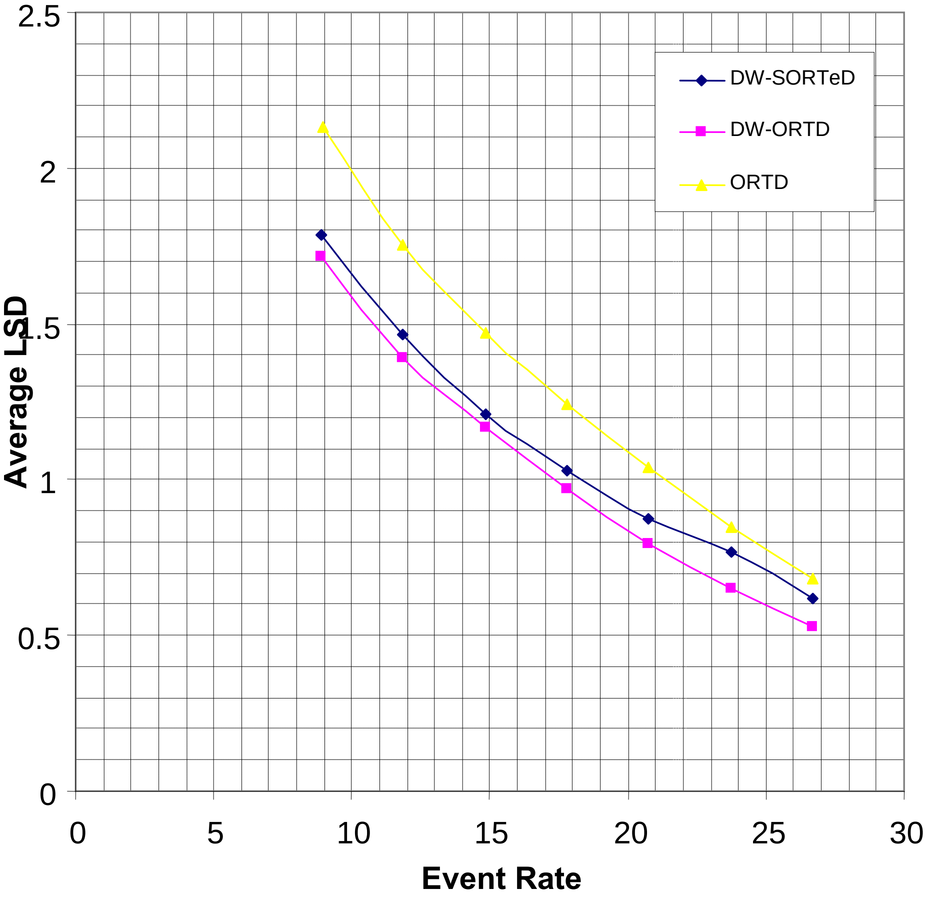


Figure 6-2. Spectral distortions (average LSD), obtained for DW-ORTD and DW-SORTeD. Modified Gardner weights were used for both. The ORTD curve (with no weighting of squared error) is also given for the comparison.

איור 6-ב. עיוותים ספקטרליים (ה-LSD הממוצע), אשר מתקבלים עבור אלגוריתמי ה-DW-ORTD ו--DW-SORTeD. משקלי Gardner מונחתים שימשו בשני האלגוריתמים. עקום של תוצאות ה-ORTD (ללא שקלול של השגיאה הריבועית) נתון גם הוא לצרכי ההשוואה.

6.2.5 הוספת האילוצים על פונקציות המאורעות ב-DW-SORTeD

במקודד המוצע אנו מקודדים את צורת פונקציות המאורעות ע"י קידוד וקטורי, כאשר ספרי הקוד אומנו על פונקציות המאורע המאולצות. בסעיף זה נראה איזו דגרדציה מכניסים האילוצים על פונקציות המאורעות. האילוצים שמופעלים על פונקציות המאורע כוללים אילוץ האי-שליליות, אילוץ המונוטוניות, אילוץ הערך המקסימלי במרכז המאורע ואילוץ ההשלמה לאחד.

בטבלה 6-ד ניתן לראות את תוצאות ההרצה של המערכת המאולצת עם הפרמטרים הטובים ביותר מול המערכת ללא האילוצים. הדגרדציה של כ-0.3 dB נצפית בין המערכת המאולצת למערכת המשתמשת בפונקציות מאורע רגעיות אופטימליות.

Table 6-d. Performance of the constrained DW-SORTeD algorithm compared to appropriate unconstrained algorithm. Both are using modified Gardner weights. The configuration of 4 events in a buffer of 15 frames is used.

טבלה 6-ד. ביצועים של האלגוריתם ה-DW-SORTeD המאולץ בהשוואה לאלגוריתם ללא אילוצים על פונקציות המאורע. שני האלגוריתמים משתמשים במשקלי Gardner מונחתים. גודל הבלוק הוא 15 מסגרות, מספר המאורעות לבלוק הוא 4.

Alg.	Last target refin.	New monot. constr.	LSD		
			Avg., [dB]	2-4dB,[%]	>4dB,[%]
Constrained DW-SORTeD	No	Yes	1.78	29.37	5.06
Unconstrained DW-SORTeD			1.47	21.32	2.15

בטבלה 6-ה מוצגות המדידות המעידות על שיפור בהתאמה אם לא משנים את וקטור המטרה האחרון בשלב של עידון וקטורי המטרה של אלגוריתם ה-DW-SORTeD (ר' סעיף 5.3.2). שיפור זה בא לידי ביטוי במערכות המשתמשות בפונקציות מאורעות מאולצות או מקוונטות. כמו כן ניתן לראות בדיקה השוואתית של שני אילוצי המונוטוניות המתוארים בסעיף 5.4.7. האילוץ הישן הוא זה שהוצע ב-[42] והאילוץ החדש הנו זה המוצג בעבודה זו בסעיף 5.4.7.

Table 6-e. . Performance of different configurations of constrained DW-SORTeD algorithm with modified Gardner weights. 4 events in a buffer of 15 frames are used.

טבלה 6-ה. ביצועים של תצורות שונות של האלגוריתם ה-DW-SORTeD המאולץ עם משקלי Gardner. גודל הבלוק הוא 15 מסגרות, מספר המאורעות לבלוק הוא 4.

Alg.	Last target refin.	New monot. constr.	LSD		
			Avg., [dB]	2-4dB,[%]	>4dB,[%]
Constrained DW-SORTeD	No	Yes	1.78	29.37	5.06
	No	No	1.79	29.79	5.09
	Yes	Yes	1.85	28.95	5.48
	Yes	No	1.86	29.88	5.7
Unconstrained DW-SORTeD			1.47	21.32	2.15

6.2.6 בחינת הביצועים כתלות במספר האיטרציות ב-DW-SORTeD

בגרף שבאיור 6-גניתן לראות את ההשפעה של מספר האיטרציות של האלגוריתם (אשר כוללות את שלב מציאת הסגמנטציה ושלב עידון וקטורי המטרה) על הביצועים של ה-DW-SORTeD המאולץ והבלתי-מאולץ. ניתן להסיק מהגרף, כי שני האלגוריתמים מתכנסים כבר באיטרציה השנייה או השלישית, כאשר עבור הפתרון המאולץ שיפור מהאיטרציה הראשונה לערך ההתכנסות הוא מזערי

(0.02 dB) וכנראה לא מצדיק את המאמץ החישובי הכרוך באיטרציות הבאות (בכל מקרה אין צורך ביותר משתי איטרציות). הסיבה לשיפור קטן כ״כ מאיטרציה לאיטרציה בפתרון המאולץ היא בכך, שאין יותר הבטחה שיהיה שיפור מאיטרציה לאיטרציה (בגלל אילוץ המונוטוניות). ע״מ לא לקלקל את הביצועים מאיטרציה לאיטרציה, קיים במימוש מנגנון של ״יציאה מוקדמת״, אשר שומר תמיד את הפתרון המיטבי ומפסיק את האלגוריתם, כאשר התוצאה מתחילה להתקלקל, במקום להשתפר. העובדה שאין כמעט שינוי מאיטרציה לאיטרציה מעידה על כך, שמנגנון זה מופעל לעתים קרובות או, לחילופין, אילוצים קשיחים על צורת פונקציות המאורע מביאות לכך, שהפתרון עבור בלוקים מסוימים מתייצב כבר באיטרציה הראשונה או השנייה.

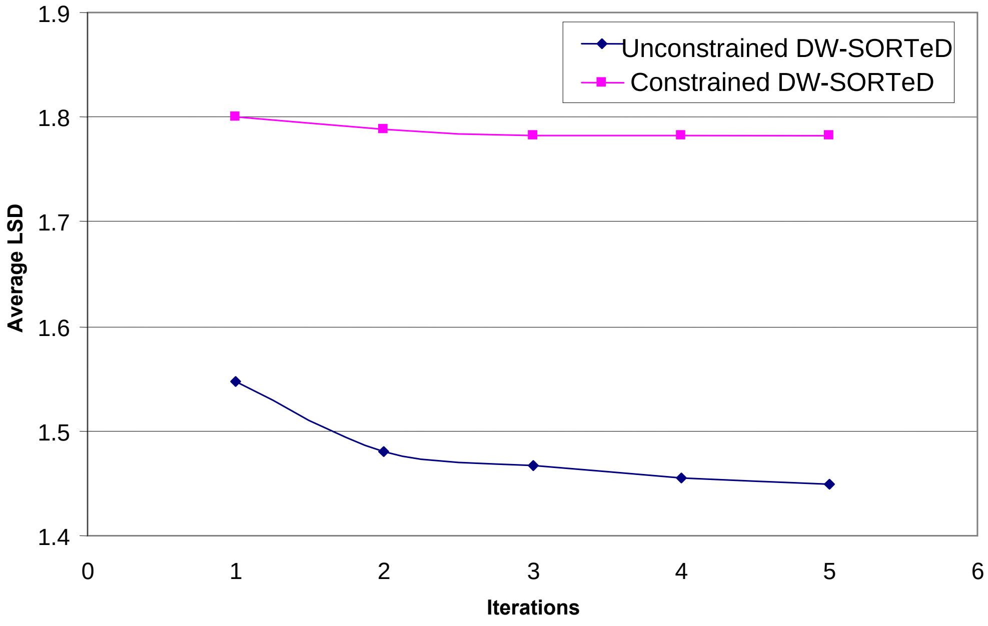


Figure 6-3. Improvement of DW-SORTeD performance vs. number of its iterations

איור 6-ג. שיפור של הביצועים של DW-SORTeD בתלות במספר האיטרציות שלו.

6.3 הערכת הביצועים של המודל עם קוונטיזציה

סדרת האימון כללה 300,000 וקטורי LSF של דוברים שונים מתוך דיבור ללא רעשי רקע של מספר גברים ונשים שונים. ספרי הקוד הנבדקים הם מסוג Split-VQ [13], כאשר מחלקים את וקטור ה-LSF לשני תתי-וקטורים של ארבעת ה-LSF התחתונים (מתאימים לתדרים הנמוכים) וששת ה-LSF העליונים (מתאימים לתדרים גבוהים). ספרי הקוד מתוכננים באופן בלתי תלוי עם מדד שגיאה ריבועית ממוצעת משוקללת. העיוות הנבדק בכל הניסויים בין שני וקטורי מעטפת ספקטרלית נעשה בעזרת LSD כפי שהוצג בסעיף 2.4.1.

6.3.1 בחינת ספר קוד לוקטורי המטרה

ספרי הקוד עבור וקטורי המטרה נוצרו ע"י אימון של כלל וקטורי ה-LSF ולא רק של וקטורי המטרה. הביצועים של ספרי הקוד שנוצרו ע"י אימון של וקטורי המטרה בלבד היו ירודים בכ-0.1 dB. הסיבה לכך היא שהפילוג של וקטורי המטרה דומה לזה של כלל וקטורי ה-LSF ואימון עם סדרת אימון הקטנה פי ארבעה עלול לקלקל את הביצועים. בנוסף לכך, נעשתה השוואה בין משקלי Gardner ומשקלי Paliwall-Atal, כאשר משתמשים בעיות WMSE בתכנון ספר הקוד ובפעולת הקוונטיזציה. כצפוי, משקלי Gardner נתנו ביצועים טובים יותר ממשקלי Paliwall-Atal, כי הם נותנים את הקירוב המיטבי של קריטריון ה-WMSE ל-LSD בהנחת שגיאות קוונטיזציה קטנות. את התוצאות ניתן לראות בטבלה 6-ו.

החלטנו להשתמש בספר הקוד הבנוי על משקלי Gardner גם עקב הביצועים המשופרים שלו וגם מכיוון שבאלגוריתם ה-DW-SORTeD אנו משתמשים במשקלי Gardner מונחתים, וכך אין צורך לחשב 2 סוגי משקלים לכל מסגרת דיבור (המעבר ממשקלי Gardner למשקלי Gardner המונחתים הוא ע"י הכפלות בקבועים בלבד).

Table 6-f. Performance of different codebooks for LSF quantization by Split-VQ. The LSF vector are split to the lowest 4 components and the highest 6 components.

טבלה 6-ו. ביצועים של ספרי קוד שונים בקוונטיזציה של וקטורי ה-LSF בסכימה של Split-VQ. פיצול הוקטורים עבור Split-VQ: 4 הרכיבים התחתונים ו-6 הרכיבים הגבוהים.

WMSE Weights	Train. data	Split-VQ, [bit]		LSD		
		Low	Hi	Avg.,[dB]	2-4dB,[%]	>4dB,[%]
Gardner	All LSFs	11	11	0.978	2.328	0
Paliwall-Atal	All LSFs			1.043	5.193	0
Gardner	All LSFs	10	10	1.127	5.480	0
Paliwall-Atal	All LSFs			1.214	9.491	0.036

6.3.2 בחינת הקידוד של פונקציות המאורע

בטבלה 6-ז ניתן לראות את התוצאות של הרצות של DW-SORTeD, כאשר מקוונטים רק פונקציות המאורע. ניתן להיווכח, כי ההבדל בין קוונטיזציה של צורת פונקציות המאורע ב-4 סיביות בלבד לבין הפתרון המאולץ ללא קוונטיזציה (ר' טבלה 6-ה) הינו 0.07dB בלבד (וזאת בין השאר עקב השילוב של קוונטיזציה לתוך תהליך של חיפוש הסגמנטציה (ר' סעיף 5.5).

Table 6-g. . Performance of DW-SORTeD with unquantized target vectors

טבלה 6-ז. הביצועים של ה-DW-SORTeD ללא קוונטיזציה של וקטורי המטרה.

Target Quant.	Event func. Quant.	LSD		
		Avg.,[dB]	2-4dB,[%]	>4dB,[%]
Unquant.	4	1.8533	30.7168	5.6272
	3	1.8767	31.3262	5.9498
	2	1.9412	32.569	6.8434

6.3.3 בחינת הביצועים של DW-SORTeD בקצבים שונים ובהשהיות שונות

מטרת ניסוי זה היא למצוא נקודות עבודה טובות לדחיסת פרמטרי ה-LSF ע"י DW-SORTeD. הוחלט למצוא נקודות עבודה מיטביות עבור מגוון קצבים בתחום של 300-380 סיביות לשנייה ועבור השהיות אלגוריתמיות של N=7,11,15 מסגרות במקודד. בטבלה טבלה 6-ח ובגרפים שבאיור 6-ד ו-באיור 6-ה מוצגים ביצועי המקודדים (של המעטפת הספקטרלית) בנקודות העבודה המועדפות עבור כ"א מההשהיות הנ"ל.

ערכי ה-LSD הממוצעים המתקבלים הינם בתחום 2.1-2.2 dB. ערכים אלה אופייניים למערכות המקדישות הפועלות בקצב של כ-490 סיביות לשנייה לקידוד המעטפת הספקטרלית [29,30] (2-2.4 dB עבור האלגוריתמים השונים, כאשר הביצועים המיטביים מושגים ע"י אלגוריתם ה-TSQ, אשר נידון בסעיף 3.4 /2.1 dB - 2.01/). כפי שניתן להיווכח מהטבלה 6-ח, השימוש ב-DW-SORTeD מאפשר להגיע לאיכויות דומות ב-300-380 סיביות לשנייה. יש לציין, כי מספר החריגים ב-DW-SORTeD הפועל ב-370 bps הוא גדול משמעותית, בהשוואה ל-TSQ הפועל בקצב של כ-480 סיביות לשנייה (6.5% של החריגים מעל 4 dB ב-DW-SORTeD לעומת 0.9% ב-TSQ, כפי שדווח ב-[30]), אך החריגות האלה ממוסכות ע"י תכונת החלקות של הספקטרום המתקבל במוצא של ה-DW-SORTeD.

מדד ה-LSD בלבד אינו מצליח להעריך את התכונות הדינמיות של ההתאמה הספקטרלית [25,56], כיוון שמדד זה מעריך את איכות ההתאמה בכל מסגרת דיבור בנפרד. ידוע, למשל, כי השתנות חלקה של הספקטרום מהווה גורם מכריע בהערכת האיכות הסובייקטיבית [25,56], אך מדד ה-LSD אינו יכול לכמת את התכונה הזאת. חלקות ההשתנות של הספקטרום היא התכונה המובנית של פעולת ה-TD, ושימוש בפונקציות מאורע מאולצות/ מקוונטות מבטיח זאת בוודאות. לעיתים, האוזן לא תבחין בחריגות נקודתיות מהספקטרום, ובלבד שההשתנות תהיה איטית ממסגרת למסגרת. מהסיבות הללו, אחוזי החריגות מהספקטרום כתוצאה מדחיסה באמצעות ה-DW-SORTeD אינן בהכרח מעידות על קיום של מספר רב של הפרעות נקודתיות, הנתפשות ע"י האוזן.

ע״מ לאמוד את איכות ההתאמה האמיתית, מקודדי המעטפת הספקטרלית מסוג DW-SORTeD שולבו יחד עם אות העירור של תקן MELP-2400 (מתקבל כך מקודד בקצב כולל של 1500-1600 סיביות לשנייה) ואות המוצא הוערך ע״י מדד ה-PESQ התקני [60]. ציונים אלו התווספו גם הם לטבלה 6-ח. גרף של ציוני ה-PESQ הממוצעים בתלות בקצב הסיביות למעטפת הספקטרלית ניתן גם בנפרד באיור 6-ה. את הציונים הנ״ל ניתן להשוות לציון של דחיסת המעטפת הספקטרית התקנית של MELP-2400 ב-1111 סיביות לשנייה (ר׳ טבלה 6-ח). יש לציין, כי סטיית התקן של המדידות על פני 20 המשפטים היא כ-0.21 והיא דומה לזו של מקודד MELP-2400 המלא (0.22)[2].

מהמדידות הנ״ל נובע, כי ניתן לחסוך כ-70% מקצב הסיביות לשידור המעטפת הספקטרלית (בהשוואה ל-MELP התקני), כאשר נגרמת כתוצאה מכך דגרדציה של 1.1–1.25 dB ב-LSD וכ-0.2 בציון ה-PESQ. יש להוסיף, כי במבחני השמיעה בלתי רשמיים נמצא כי ההבדלים לעומת קידוד הספקטרום התקני (MELP-2400) מורגש קלות בלבד, והאיכות המתקבלת קרובה לזו של מקודד ה-MELP התקני (נמצא, כי ההבדלים בין אותות המוצא של שני המודלים כמעט ולא נבחנים ב-19 משפטים מתוך 20).

ניתן לראות בבירור את הדגרדציה של הביצועים במעבר מאורך בלוק של 11 מסגרות לאורך בלוק של 7 מסגרות, אך אין דגרדציה משמעותית במעבר מאורך בלוק של 15 לאורך בלוק של 11 מסגרות. אי לכך, נציע להשתמש בהשהיה אלגוריתמית של 11 מסגרות (247.5 מילישניות). תצורה זו תאפשר להוציא את המיטב מהמודל ה-SORTeD עם השהיה קטנה יותר (השיפור שבין תצורת ה-11 מסגרות לתצורה של 15 מסגרות אינו מצדיק את הגדלת ההשהיה ב-90 מילישניות נוספות). עם מעוניינים בהשהיה קטנה יותר, אפשר להשתמש בתצורה שמשתמשת בבלוק באורך 7 מסגרות (השהיה אלגוריתמית של 157.5 מילישניות). עם ההשהיה המוקטנת ניתן לקבל ביצועים דומים למערכת עם N=11 ע״י הגדלה של קצב השידור של המעטפת הספקטרלית בכ-80 סיביות.

מהתבוננות בגרפים באיור 6-ד ובאיור 6-ה ניתן להסיק, כי עבור שינויים קטנים השינוי ב-LSD לא תמיד עוקב אחר השינוי ב-PESQ – זה מאמת את העובדה שמדד ה-LSD לא רגיש להתנהגות הדינמית, להבדיל מהמנגנון האנושי, הממודל ע״י ה-PESQ.

[2] יש לציין, כי ממוצע PESQ של המשפטים הנאמרים ע״י נשים קטן בכ-0.1 בהשוואה לגברים. הדבר תקף גם לגבי מקודד MELP-2400 התקני.

Table 6-h. The evaluation of DW-SORTeD for spectral envelope coding. The "best for a given rate and block length" setups are given in a 300-380 bps range for spectral coding envelope and block lengths of 7,11 and 15 frames. The standard deviation of PESQ measurements is about 0.21.

טבלה 6-ח. הערכת הביצועים של מקודד ה-DW-SORTeD עבור המעטפת הספקטרלית של הדיבור. נתונות התצורות הטובות ביותר עבור הקצבים הרצויים (בתחום 300 -380 סיביות לשנייה לקידוד המעטפת הספקטרלית) ואורך בלוק של 7, 11 ו-15 מסגרות. סטיית התקן של מדידות ה-PESQ היא כ-0.21.

M/N	Split-VQ cdbk [bit]		Ev. func. shape cdbk [bit]	Ev. func. Len. [bit]	Sp.Env. Rate, [Bps]	LSD			PESQ
	Low	Hi				Avg., [dB]	2-4dB, [%]	>4dB, [%]	
4/15	12	12	4	3	367	2.09	38.30	6.56	2.82
	12	12	2	3	343	2.16	38.98	7.70	2.81
	11	9	3	3	308	2.20	42.28	6.99	2.78
3/11	12	12	4	3	376	2.08	36.80	7.45	2.83
	11	9	4	3	327	2.17	39.23	7.20	2.79
	10	8	4	3	303	2.24	41.28	8.56	2.78
2/7	12	12	3	3	381	2.14	35.99	9.57	2.77
	11	9	4	3	343	2.19	38.03	8.96	2.76
	10	8	3	3	305	2.27	41.18	9.53	2.73
MELP-2400					1111	0.94	1.50	0.00	2.99

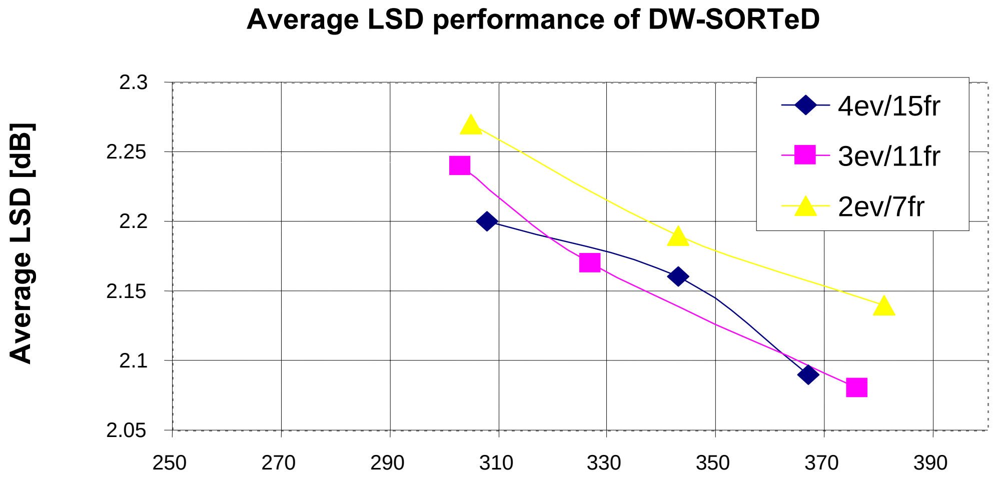


Figure 6-4. The evaluation of DW-SORTeD for spectral envelope coding by average LSD scores. The "best for a given rate and block length" setups are given in a 300-380 bps range for spectral coding envelope and block lengths of 7,11 and 15 frames.

איור 6-ד. הערכת הביצועים של מקודד ה-DW-SORTeD עבור המעטפת הספקטרלית של הדיבור ע"י מדד ה-LSD הממוצע. נתונים התצורות הטובות ביותר עבור הקצבים הרצויים (בתחום 300 -380 סיביות לשנייה לקידוד המעטפת הספקטרלית) ואורך בלוק של 7, 11 ו-15 מסגרות.

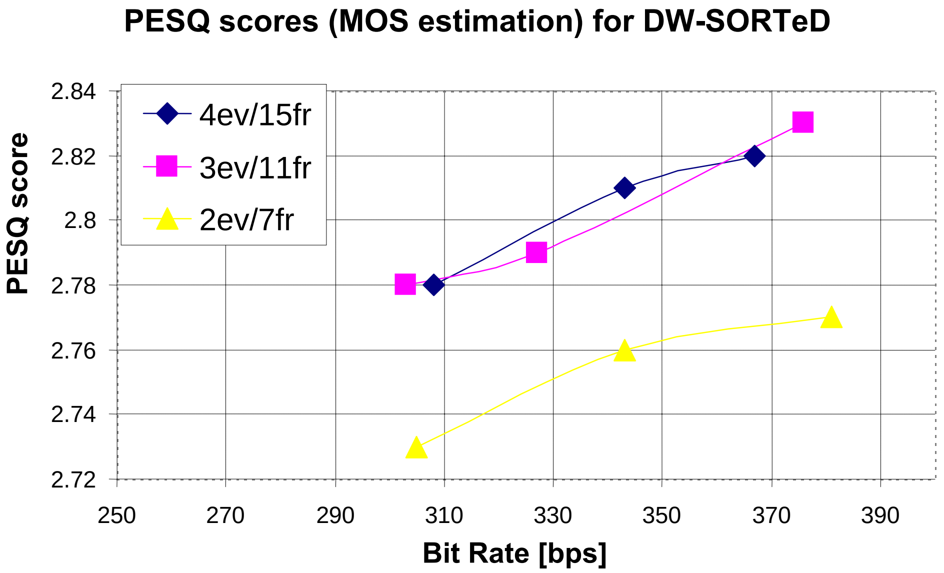


Figure 6-5. The evaluation of DW-SORTeD for spectral envelope coding, combined with the MELP standard excitation, using PESQ scores. The "best for a given rate and block length" setups are given in a 300-380 bps range for spectral envelope coding and block lengths of 7, 11 and 15 frames.

איור 6-ה. הערכת הביצועים של מקודד ה-DW-SORTeD עבור המעטפת הספקטרלית של הדיבור בשילוב עם קידוד העירור התקני של תקן MELP-2400 באמצעות ציון ה-PESQ. נתונים התצורות הטובות ביותר עבור הקצבים הרצויים (בתחום 300 -380 סיביות לשנייה לקידוד המעטפת הספקטרלית) ואורך בלוק של 7, 11 ו-15 מסגרות.

6.4 סיכום

בחנו בפרק זה את אלגוריתם DW-SORTeD המוצע לקידוד המעטפת הספקטרלית המיוצגת ע"י וקטורי LSF. הבדיקות שנעשו על המודל כללו:

1) השוואת הביצועים של המודלים ORTD, DW-ORTD, DW-SORTeD.
2) בחינת פרמטרים שונים של ה-DW-SORTeD.
3) בחינת ההתכנסות של ה-DW-SORTeD.
4) בחינת הביצועים של ה-DW-SORTeD בשילוב עם קוונטיזציה בנקודות עבודה שונות.

אלגוריתם DW-SORTeD מאפשר להגיע לקצבים של 300-380 סיביות לשנייה עם LSD ממוצע של 2.1-2.25 dB וציון PESQ של כ-2.8 (זאת בשילוב עם קידוד עירור תקני של MELP-2400). איכות ההתאמה של DW-SORTeD, הפועל בקצב של 370 סיביות לשנייה, דומה לזו המתקבלת ע"י אלגוריתם ה-TSQ [29], הפועל בקצב של 490 סיביות לשנייה. השהיה אלגוריתמית של 11 מסגרות נדרשת לקבלת ביצועים אלה, אך ניתן להוריד את ההשהיה ל-7 מסגרות בלבד ולהתקרב לביצועים אלה אם מגדילים את קצב השידור בכ-80 סיביות לשנייה נוספות.

פרק 7

דחיסת פרמטרי העירור באמצעות אלגוריתם DW-SORTeD

7.1 מבוא

פירוק ה-TD הנו כלי בעל עוצמה רבה וניתן להשתמש בו לצורך הורדת יתירות זמנית ברצף וקטורי פרמטרים כללי. בפרק זה ננצל את ה-TD ע״מ לדחוס פרמטרי עירור אחדים, כגון ערך ה-pitch והאנרגיה. פרמטרים אילו קיימים ברוב המקודדים הפרמטריים בצורה זו או אחרת, ולכן ניתן לשלב דחיסת פרמטרים אלה באמצעות ה-TD במגוון רחב של מקודדים, כגון מקודדי LPC, MELP, סינוסואידליים ועוד.

ברור כי אופי מסלולי ההשתנות של פרמטרים אילו שונים בהשוואה לפרמטרי ה-LSF. הם פחות חלקים, בעלי תחום דינמי גדול וקשה לקבוע קריטריון למינימיזציה עבורם. אמנם, ניתן לבצע את ה-TD בנפרד לאנרגיה ול-pitch, ואכן נעשו בעבר ניסיונות לעשות זאת עם מודיפיקציות שונית של TD [390,520]. למעשה, במקרה סקלרי פעולת ה-RTD היא פשוט מסגרת נוחה לביצוע אינטרפילציה לינארית במרחקים משתנים. בפרק זה נבחן ביצועי הסכימה המוצעת (DW-SORTeD) בקידוד הפרמטרים האלה כסקלרים וכן נציע ונבחן דרך לקידוד ה-pitch והאנרגיה ע״י ביצוע ה-TD לשניהם במשותף. גישה של TD לשני הפרמטרים במשותף מניבה דחיסה עמוקה במיוחד והיא ניתנת ליישום תוך שימוש במשקלות דינמיים אמפיריים.

פרק זה בנוי כדלהלן: תחילה, נבחן תכונת ה-RTD המאולץ, אשר הופכת אותו להיות מתאים לביצוע ה-TD של זוג פרמטרי העירור יחדיו. לאחר מכן, נציג מודל ספציפי לביצוע ה-TD של פרמטרי העירור במשותף ונפרט על העיבוד המקדים של וקטורי פרמטרי העירור בכניסה לשלב ה-TD ועל המשקלים הדינמיים שבשימוש ה-TD. בסעיף האחרון נציג תוצאות ניסיוניות לביצוע ה-TD לפרמטרי העירור בנפרד ובמשותף.

7.2 התאמת ה-RTD המאולץ לביצוע ה-TD לוקטורי פרמטרים מנורמלים

7.2.1 מבוא

פעולת ה-TD הינה שיטה כללית לפרמטריזציה של רצף וקטורי כללי. הזכרנו קודם לכן, כי שיטת ה-RTD עם אילוץ ההשלמה לאחד ניתנת לפירוש גאומטרי כקירוב המסלול שמיצרים וקטורי הפרמטרים במרחב ה-p-מימדי ע״י עקום ליניארי למקוטעין. כלומר, כל רצף וקטורים שניתן לקירוב טוב כליניארי למקוטעין (עם אורך קטע ממוצע גדול מספיק ע״מ לקבל דחיסה טובה) ניתן למידול ע״י ה-TD. וקטורי המטרה במקרה זה מהווים נקודות השבירה של עקום זה, ווקטורים ששייכים לסגמנט מסוים מוטלים על הקו הישר המחבר בין זוג וקטורי המטרה הסובבים אותם. (נשים לב, כי אילוץ המונוטוניות מחייב בנוסף, כי ההתקדמות (עם הזמן) לאורך עקום זה תהיה תמיד בכוון אחד, כלומר אם מתקיים $n_k < m_1 < m_2 < n_{k+1}$, אזי $\left\|\hat{\mathbf{y}}(n_{k+1})-\hat{\mathbf{y}}(m_2)\right\| \le \left\|\hat{\mathbf{y}}(n_{k+1})-\hat{\mathbf{y}}(m_1)\right\|$). שימוש במשקלים דינמיים ״מקלקל״ את הדמיון הגאומטרי, אך מאפשר לשלוט בדיוק ההתאמה בנפרד לכל רכיב הווקטור ובכל רגע זמן n.

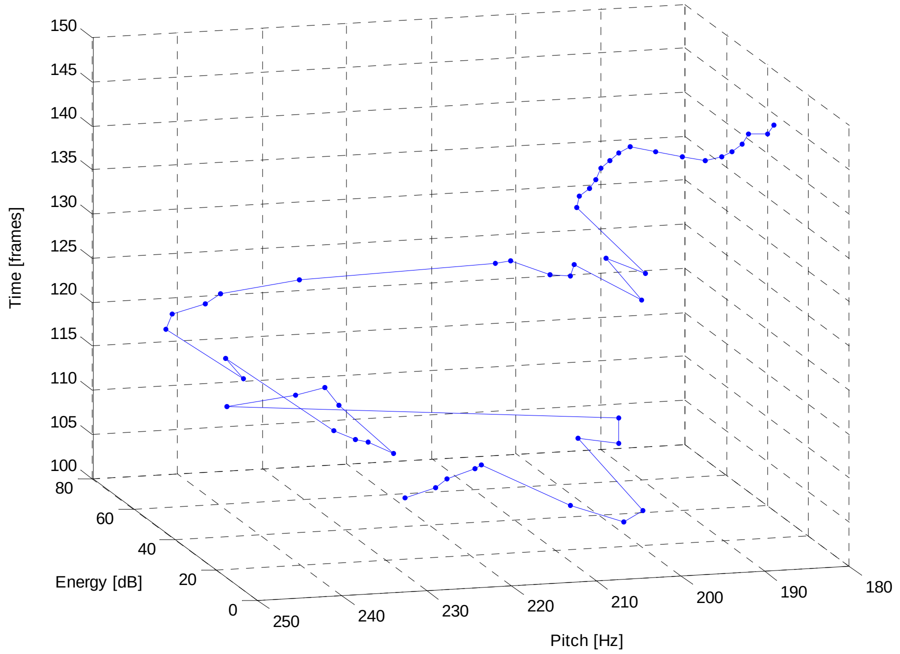


Figure 7-1. Pitch-Energy joint trajectory in time

איור 7-א. מסלול ההשתנות של האנרגיה וה-pitch לאורך זמן .

תכונות אילו של ה-TD הופכות אותו מתאים לפרמטריזציה של תדר ה-pitch ו/או האנרגיה. ניתן להפעיל פעולת ה-TD על כל אחד מפרמטרים אילו בנפרד וכן, כפי שנראה בהמשך הסעיף, במשותף (ר׳

איור 7-א להדגמת המסלול שמייצרות האנרגיה וה-pitch). נכנה את צמד ה-pitch והאנרגיה המיועד ל-TD *וקטור פרמטרי העירור*, בדומה לוקטורי הפרמטרים הספקטרליים ששימשו כוקטורי הכניסה למערכת ה-TD עבור המעטפת הספקטרלית של הדיבור. בפרק זה נתייחס לפרמטר ה-pitch של מסגרת דיבור בנפרד מההחלטה על סוג המסגרת (קולית או א-קולית), כלומר נניח כי אינפורמציה זו תקודד ותשודר בנפרד, וערך של פרמטר ה-pitch במסגרת א-קולית הנו שרירותי (שימוש במשקלים דינמיים לביצוע ה-TD יאפשר לנטרל את השפעת ה"טעות" ב-pitch במסגרות א-קוליות על השגיאה הכוללת בעת ביצוע ה-TD).

7.2.2 RTD עם אילוץ השלמה לאחד תחת התמרה אפינית (affine transform)

נבחן, כעת, איך ניתן להשתמש במידול ה-TD של ה-pitch והאנרגיה במשותף. פרמטרים אילו הם בעלי טווח ערכים שונה לחלוטין, ונראה לכאורה כי לא ניתן לעשות זאת. על הרכיבים של וקטור להיות בעלי טווח ערכים דומה, כדי למנוע מצבים של משואות סינגולריות ולאפשר שקלול יחסי בין פרמטרים אילו במידת הצורך (ע"מ להגביר רגישות לשגיאות בפרמטר אחד ע"ח פרמטר שני בזמנים מסוימים). התכונה הבאה של ה-*RTD עם אילוץ ההשלמה לאחד* פותחת את האפשרות לנרמול של כ"א מרכיבי וקטור הפרמטרים לתחום הרצוי:

למה 1: יהי $\hat{\mathbf{y}}(n) = \mathbf{a}_k\phi_k(n) + \mathbf{a}_{k+1}\phi_{k+1}(n),\ \ n_k \le n < n_{k+1}$ פירוק ה-RTD עם האילוץ של $\phi_{k+1}(n) + \phi_k(n) = 1$ של רצף וקטורי המטרה $\mathbf{y}(n)$, $1 \le n \le N$ והשגיאות הרגעיות ניתנות ע"י $E(n) = \sum_{i=1}^{p} w_i(\mathbf{y}(n) - \hat{\mathbf{y}}(n))^2$. תהיה $\tilde{\mathbf{y}}(n) = \mathbf{C}\mathbf{y}(n) + \mathbf{d}$ טרנספורמציה אפינית כלשהי של וקטורי הכניסה, אזי פירוק ה-RTD של הרצף $\tilde{\mathbf{y}}(n)$, $1 \le n \le N$ יהיה: $\hat{\tilde{\mathbf{y}}}(n) = (\mathbf{C}\mathbf{a}_k + \mathbf{d})\phi_k(n) + (\mathbf{C}\mathbf{a}_{k+1} + \mathbf{d})\phi_{k+1}(n),\ \ n_k \le n < n_{k+1}$ ויביא לשגיאה רגעית $E(n) = (\mathbf{y}(n) - \hat{\mathbf{y}}(n))^T \mathbf{C}^T \mathbf{W}(n) \mathbf{C}(\mathbf{y}(n) - \hat{\mathbf{y}}(n))$, כאשר $\mathbf{W}(n)$ היא מטריצת משקלות אלכסונית.

<u>הוכחה:</u>

נפעיל טרנספורמציה אפינית על $\hat{\mathbf{y}}(n)$:

$$\mathbf{C}\hat{\mathbf{y}}(n) + \mathbf{d} = \mathbf{C}(\mathbf{a}_k\phi_k(n) + \mathbf{a}_{k+1}\phi_{k+1}(n)) + \mathbf{d} = \mathbf{C}(\mathbf{a}_k\phi_k(n) + \mathbf{a}_{k+1}\phi_{k+1}(n)) + \mathbf{d}(\phi_k(n) + \phi_{k+1}(n)) =$$
$$= (\mathbf{C}\mathbf{a}_k + \mathbf{d})\phi_k(n) + (\mathbf{C}\mathbf{a}_{k+1} + \mathbf{d})\phi_{k+1}(n).$$

הש

תמשנו כאן בתכונת ההשלמה לאחד של פונקציות המאורעות ($\phi_{k+1}(n) + \phi_k(n) = 1$).

השגיאה הרגעית:

$$E(n) = (\mathbf{y}(n) - \hat{\mathbf{y}}(n))^T \mathbf{W}(n)(\mathbf{y}(n) - \hat{\mathbf{y}}(n)) =$$
$$= (\mathbf{C}\mathbf{y}(n) + \mathbf{d} - \mathbf{C}\hat{\mathbf{y}}(n) - \mathbf{d})^T \mathbf{W}(n)(\mathbf{y}(n) + \mathbf{d} - \mathbf{C}\hat{\mathbf{y}}(n) - \mathbf{d}) =$$
$$= (\mathbf{y}(n) - \hat{\mathbf{y}}(n))^T \mathbf{C}^T \mathbf{W}(n) \mathbf{C}(\mathbf{y}(n) - \hat{\mathbf{y}}(n)).$$

□

בעזרת בלמה 1, נציע שיטה לביצוע ה-TD עבור האנרגיה (ב-dB) ותדר ה-pitch (בהרצים) במשותף. לכל בלוק אנליזה, ננרמל את הפרמטרים (ע"י הכפלה בקבוע והוספת הסט קבוע, כלומר, המטריצה **C** תהיה אלכסונית), כך שימצאו בטווח ערכים דומה זה לזה. לאחר מכן נבצע פעולת ה-RTD באמצעות האלגוריתמים DW-SORTeD או DW-ORTD עם אילוץ ההשלמה לאחד. לסיום, נמיר את וקטורי המטרה שהתקבלו לטווח המקורי שלהם ע"י פעולה לינארית הפכית לזו שנעשתה בשלב הנרמול. בסעיפים הבאים נפרט על הדרך בה מנרמלים את פרמטרי העירור ועל המשקלים הדינמיים ששימשו ל-DW-ORTD.

7.3 מודל לביצוע ה-TD למספר פרמטרי העירור

7.3.1 תיאור כללי

מודל ה-TD של וקטור פרמטרי העירור יתבסס על DW-ORTD מאולץ (ר' סעיף 5.4.7) עם נרמול מקדים של פרמטרי העירור (ר' איור 7-ב). נשים לב, כי פעולת הנרמול נעשית במקודד לצורך ביצוע ה-TD, ושום מידע צד לא מועבר לצד המפענח. אם וקטור פרמטרי העירור הוא במימד אחד (כלומר מתנוון לסקלר), אין צורך בביצוע הנרמול.

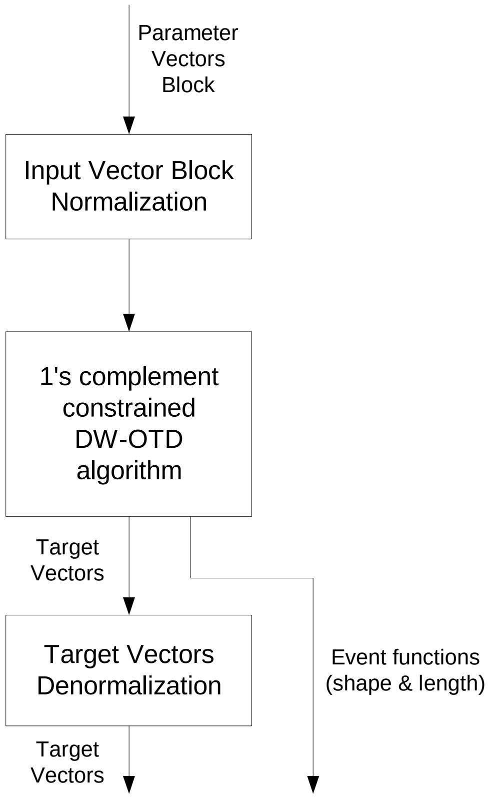


Figure 7-2. Excitation TD

איור 7-ב פעולת ה-TD לוקטורי פרמטרי עירור

בהמשך הפרק נפרט על השלבים באלגוריתם, תוך הדגשת השוני בין אלגוריתם זה לפעולת ה-TD של פרמטרי ה-LSF (המשמשות לקידוד המעטפת הספקטרלית).

7.3.2 עיבוד מקדים של פרמטרי העירור

ע״מ שביצוע של TD של אנרגיה ו-pitch במשותף יצליח, חשוב, כי האזורים הא-קוליים יסומנו כאזורים בהם ה-pitch יכול לקבל ערך כלשהו. חשוב לכן, את האזורים האלו לסמן כא-קוליים. ישנם תקנים (כמו MELP-2400, ר׳ סעיף 2.5) המציינים את קיום ה-pitch גם באזורים א-קוליים בפועל (״טעויות״ אלו מפוצות לאחר מכן, ע״י המידע על מידת הקוליות של המסגרת שנמצא בה pitch). במקרים כאלו, יש לסווג מסגרות כא-קוליות ע״מ להגדיל את אחוז המסגרות הא-קוליות לשיפור ההתאמה (באזורים הא-קוליים שגיאות ב-pitch אינן תורמות לשגיאה הכוללת, ולכן ההתאמה תיעשה לאנרגיה בלבד). אנו בחרנו להשתמש באלגוריתם [53] לקביעה באם מסגרת היא קולית או א-קולית. שיטה זו משתמשת בספי אנרגיה ומדידות האוטוקורלציה, לביצוע הסיווג. ניתן כמובן להשתמש בכל שיטה אחרת, כולל האינפורמציה הטמונה במדידות הקוליות של תקן ה-MELP עצמו. באזורים המסווגים כא-קוליים, נקבע ערך ה-pitch ע״י אינטרפולציה לינארית של המסגרות הקוליות

הסובבות אותם, ע״מ לקבל מסלול ההשתנות של ה-pitch רציף. פעולות אילו נעשות כשלב מקדים לאלגוריתם המתואר באיור 7-ב.

7.3.3 נרמול

האנרגיה נמדדת ב-dB והיא בעלת טווח ערכים אפשריים מוגדר (למשל, הערכים המקודדים בתקן MELP נעים בין 10 ל-77 dB). טווח אפשרי של תדר ה-pitch הנו בין 50 ל-400 Hz. טווחים של פרמטרים אלו ב-N מסגרות עוקבות (כאורך הבלוק ב-TD) יהיו צרים יותר ומשתנים לאורך זמן.

מטרת הנרמול, כאמור, להביא את כל הרכיבים של וקטורי המטרה לטווח ערכים אחיד (לדוגמה, הנע בין אפס לאחד). נרמול קבוע (בהתאם לספים ידועים) עלול להביא לכך כי שהטווחים האפקטיביים יהיו שונים. זאת מכיוון שטווח הערכים בבלוק האנליזה (עליו מופעל אלגוריתם ה-SORTeD) עלול להיות קטן מהטווח המקסילי האפשרי. ע״מ לקרב את טווחי הפרמטרים בצורה טובה יותר, נרמול זה ייעשה לכל בלוק אנליזה של וקטורי כניסה (המיועד לפירוק ה-TD) בנפרד. אולם, בלוקים עוקבים אינם בלתי תלויים עקב החפיפה של חלק מוקטורי הכניסה ווקטורי המטרה, דבר שמחייב כי פרמטרי הנרמול בבלוקים עוקבים יהיו דומים זה לזה, לעיתים על חשבון זה, שהטווחים של הרכיבים של וקטורי העירור יהיו דומים , אך לא שוים.

בפועל, ע״מ לבצע נרמול של צמד פרמטרי העירור – אנרגיה ותדר ה-pitch, נבצע עקיבה של ה-pitch הממוצע של בלוק וכן האנרגיה המקסימלית והמינימלית של בלוק. העקיבה נעשית עם מקדמי שכיחה אדפטיביים, הנקבעים בצורה אמפירית. ההחלקה נעשית בכדי למנוע השתנויות חדות של פרמטרי הנרמול מבלוק לבלוק. השתנויות מהירות של קבועי הנרמול עלולות לפגוע בתנאי הקצה של בלוק אנליזה, שכן מאורע ראשון של הבלוק הנוכחי הוא בעצם המאורע האחרון של הבלוק הקודם. באיור 7-ג ואיור 7-ד ניתן לראות את העקיבה אחר תדר ה-pitch והאנרגיה, המשמשת לחישוב קבועי הנרמול.

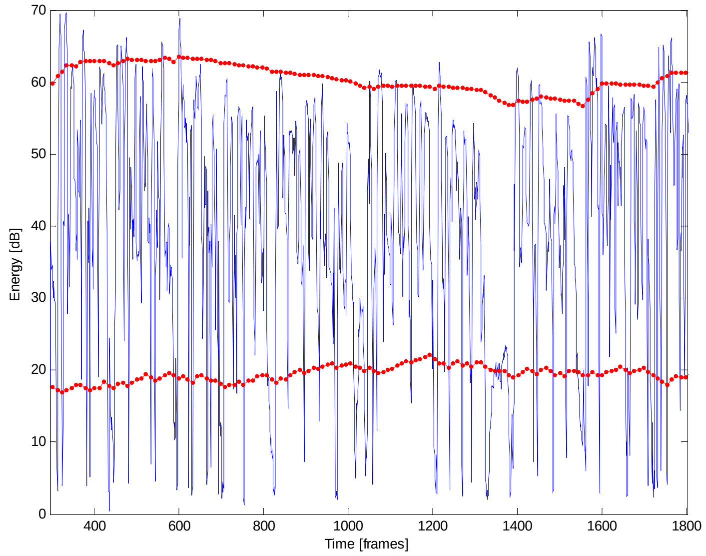


Figure 7-3. Minimum and maximum energy tracking for excitation parameter vector normalization

איור 7-ג. עקיבה אחר הערך המכסימלי והמינימלי של האנרגיה, המשמשת לשם נורמליזציה של וקטור פרמטרי העירור

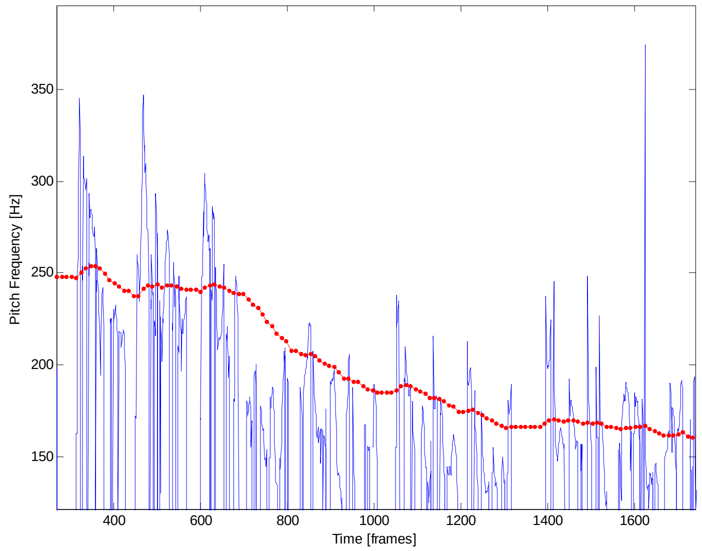


Figure 7-4. Average pitch frequency tracking for excitation parameter vector normalization

איור 7-ד. עקיבה אחר הערך הממוצע של תדר ה-pitch, המשמשת לנורמליזציה של וקטור פרמטרי העירור

בהינתן ה-pitch הממוצע P_{avg} ואנרגיה מינימלית $(G_{\min})$ ומקסימלית $(G_{\max})$, פעולת הנרמול תינתן ע״י:

$$\begin{pmatrix} \tilde{P}(n) \\ \tilde{G}(n) \end{pmatrix} = \begin{pmatrix} 1/P_{avg} & 0 \\ 0 & 1/(G_{\max}-G_{\min}) \end{pmatrix} \begin{pmatrix} P(n) \\ G(n) \end{pmatrix} + \begin{pmatrix} 0 \\ -G_{\min}/(G_{\max}-G_{\min}) \end{pmatrix} \quad (5.21)$$

נשים לב, כי הנרמול של ה-pitch נעשה ע״י הערך הממוצע ולא ערך המקסימום, מחשש לתדר pitch כפול, שעלול לגרום לאי-יציבות של מדידת המקסימום. יש לציין כי ערכי הקיצון והממוצעים הנזכרים לעיל הם רק קירוב לערכים האמיתיים עקב ההחלקה היחידות של האנרגיה הם דציבלים, ושל ה-pitch – הרצים.

7.3.4 קביעת המשקל הדינמי לביצוע ה-DW-ORTD

בעיית קביעת משקלים לקריטריון השגיאה לביצוע TD מתקשרת למדד האיכות של ייצוג פרמטרי העירור (כפי שזה מתבטא באות המוצא). בהעדר מודלים ברורים, נבחר במדד של שגיאה ריבועית ממוצעת יחד עם אחוז החריגות הגדולות מהממוצע. בכ״א מפרמטרי העירור, חריגה גדולה נקודתית עלולה להביא לדגרדציה סובייקטיבית גדולה, עקב הכנסת עיוותים הנשמעים באות המוצא. לדוגמא, העלאה נקודתית חדה באנרגיה, תייצר צלילים מפריעים ותפגע באיכות הסובייקטיבית, לעומת זה, טעויות קטנות שעלולות להביא לערך ה-MSE גדול יותר משגיאה נקודתית גדולה, לא יביאו לדגרדציה ניכרת באיכות הנתפשת של אות המוצא. נמצא בנוסף, כי האוזן אינה רגישה לשינויים קטנים של ה-pitch, ובלבד שיישאר חלק מספיק ולא יהיו חריגות גדולות מה-pitch המקורי.

השיקולים הנ״ל מביאים לקביעת המשקלים (ר׳ איור 7-ה ואיור 7-ו) תוך שימוש בקווים המנחים הבאים:

1) אנרגיה: משקל שגיאה גדל ככל שההפרש בין אנרגיות עוקבות גדל, וזאת בתנאי שההפרש גבוה מערך סף מסוים והאנרגיה עצמה גבוהה מערך סף מסוים. הדבר נעשה כך, כיוון שהאזורים בהם האנרגיה משתנה בצורה מהירה הם אזורים החשובים מבחינת השמיעה האנושית והשגיאות באזורים אלו יביאו לדגרדצית האיכות הסובייקטיבית גדולה יותר, מאשר שינויים קלים באזורים הסטציונריים. (דוגמה לאזורים רגישים כאלה יכולה להיות עיצור קצר בתוך רצף של מסגרות הקוליות.)

2) pitch: המשקל הוא אפס באזורים הא-קוליים ומוגבר יחסית לשאר המסגרות במסגרות הראשונות והאחרונות של הרצפים הקוליים של הבלוק. זאת ע״מ להקטין שגיאות בקצוות של האזורים הקוליים, אשר שוכנים לרצפים הא-קוליים, בהם ערך ה-pitch לא משפיע על פעולת ה-TD, ולכן ייתכנו סטיות משמעותיות מהערכים המקוריים..

3) נמצא, כי השגיאות באנרגיה הן מורגשות יותר מהשגיאות ב-pitch, לכן הוחלט להגדיל משקלי השגיאה עבור רכיב האנרגיה, לעומת רכיב ה-pitch.

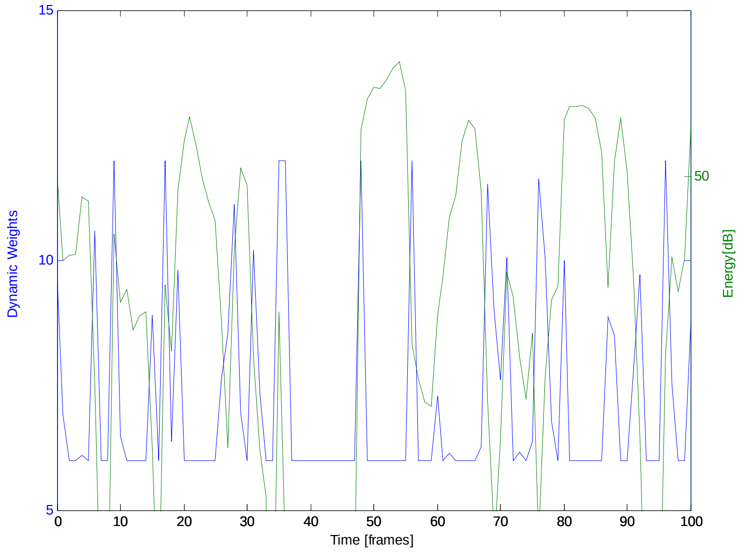


Figure 7-5. Dynamic energy weighting for joint pitch-energy quantization with DW-SORTeD algorithm. The energy is shown in green , and its weighting is displayed in blue.

איור 7-ה. משקלים דינמיים לאנרגיה עבור קידוד משותף של תדר ה-pitch והאנרגיה באמצעות DW-SORTeD. האנרגיה מסומנת בירוק וערכי המשקולת המתאימות מסומנים בכחול.

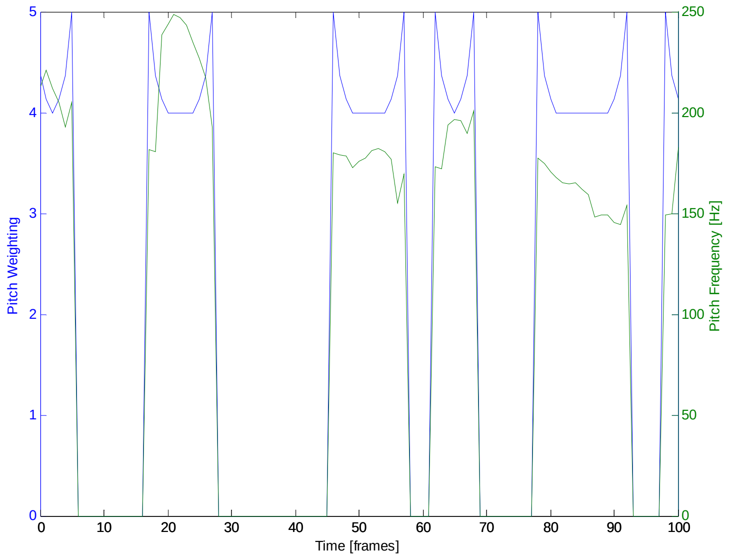


Figure 7-6. Dynamic pitch weighting for joint pitch-energy quantization with DW-SORTeD algorithm. The pitch is shown in green, and its weighting is displayed in blue.

איור 7-ו. משקלים דינמיים לתדר ה-pitch עבור קידוד משותף של תדר ה-pitch והאנרגיה באמצעות DW-SORTeD. ה-pitch מסומן בירוק וערכי המשקולת המתאימות מסומנים בכחול.

7.3.5 התאמת ה-DW-SORTeD לפרמטרי העירור

פרמטרי עירור (בייחוד אנרגיה) אינם חלקים מספיק, ולכן קצב מאורעות מצומצם ופונקציות עירור חלקות לא יקרבו את הפרמטרים היטב. עקב כך הוחלט לבטל את אילוץ המונוטוניות בעת יצירת ספרי הקוד לפונקציות המאורע. ע״מ להקטין את הסיבוכיות של החיפוש וכן לפשט את האלגוריתם, המאורע האחרון ממוקם תמיד בקצה הבלוק. במימוש האופייני של המערכת, מופקים שלושה מאורעות חדשים לבלוק (M=3) , כלומר יש לבצע חיפוש רק עבור שני המאורעות. בדרך זו מצמצמים גם את כמות הסיביות לשידור אורכי הסגמנטים (שכן יש לקודד רק שני אורכים במקום שלושה). מספיק לעשות איטרציה אחת של אלגוריתם ה-DW-SORTeD. בדומה ל-TD של פרמטרים ספקטרליים, גם כאן נשלב בין TD לקוונוט, רק שהפעם קוונוט וקטורי המטרה יהיה סקלרי (5 סיביות לאנרגיה ו-5-6 סיביות למחזור ה-pitch, שניהם בסקלה לוגריתמית).

7.4 בחינת ביצועים של קידוד פרמטרי העירור באמצעות ה-TD.

7.4.1 כללי

בסעיף זה נאמוד את הביצועים של האלגוריתם על פני 20 משפטים מתוך בסיס הנתונים TIMIT (מחציתם נאמר ע״י גברים ומחציתם ע״י נשים). במסגרת בחינת המערכת, נציג שגיאה ריבועית ממוצעת ואחוז השגיאות הגדולות עבור כ״א מפרמטרי העירור (pitch ואנרגיה). בנוסף לבדיקת המערכת שמבצעת TD לשני פרמטרי עירור במשותף, נבחן גם מערכת, שמבצעת TD לכ״א מפרמטרי העירור בנפרד.

לקראת ביצוע ה-TD, פרמטרי הכניסה שונו במקצת. אנרגית האות נמדדת ב-MELP התקני פעמיים לאורך כל מסגרת. אנו נבחר רק ערך אחד לצורך הקוונטיזציה. הערך השני לא ישודר. ערכי ה-pitch נלקחים אחרי שאלגוריתם הסיווג [53] הופעל, וחלק מהמסגרות שנקבעו ע״י MELP כקוליות הפכו לא-קוליות. מדידות ה-MSE ייעשו בין פרמטרי MELP המקוונטים ופרמטרים שעברו קוונוט ו-TD במשותף.

7.4.2 דוגמת הרצה

ניתן לראות באיור 7-ז את פרמטרי העירור לאחר ה-TD בהשוואה לפרמטרים, המקוונטים סקלרית. וקטורי המטרה מקוונטים כאן סקלרית, 5 סיביות לכל רכיב, וצורות פונקציות המאורעות קודדו ב-2 סיביות בלבד. דחיסה זו הניבה חסכון של 40% בקצב השידור בהשוואה לקוונטיזציה סקלרית, בהתחשב במידע צד של סיווג קולי/א-קולי שיש לשדר בנוסף לוקטורי המטרה ופונקציות המאורעות. ההתאמה נעשתה באמצעות 4 מאורעות בבלוקים של 11 מסגרות.

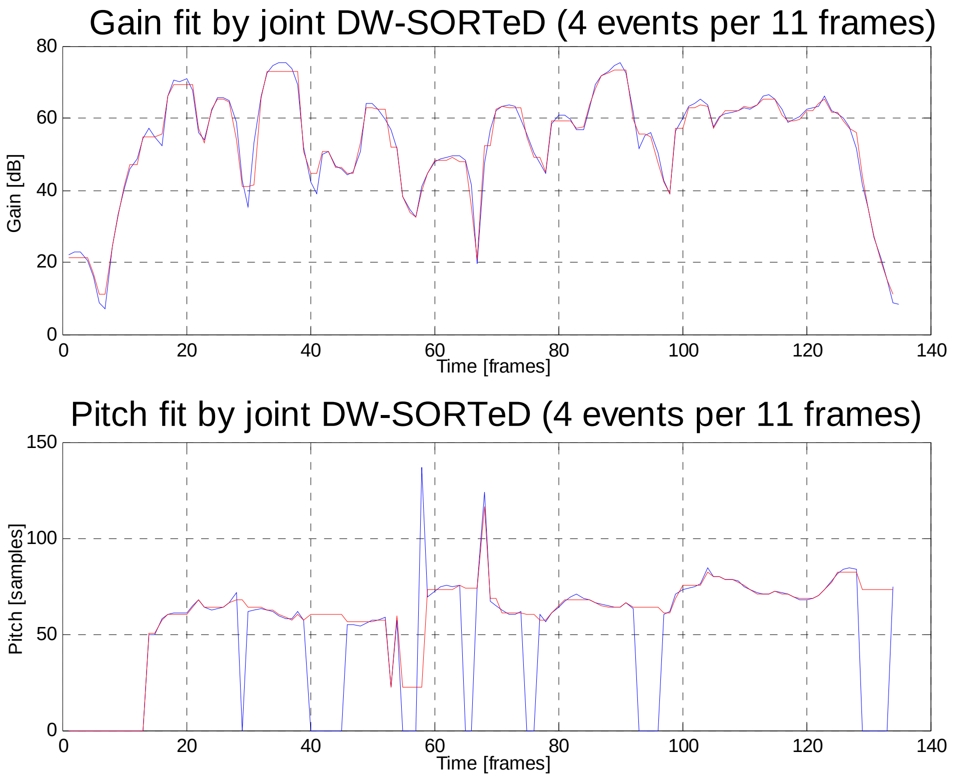


Figure 7-7. Joint DW-SORTeD of pitch and energy. The run is performed with 4 events in a block of 11 frame length. Scalar quantized vectors are given in blue, and the results of DW-SORTeD quantization are given in red.

איור 7-ז. הפעלת ה-DW-SORTeD על פרמטרי ה-pitch והאנרגיה במשותף. ההרצה נעשתה עם 4 מאורעות בבלוק של 11 מסגרות. פרמטרי העירור לאחר קוונטיזציה סקלרית נראים בכחול ומוצא הקוונטיזציה באמצעות DW-SORTeD נראים באדום.

בטבלה 7-א מוצגות תוצאות כמותיות של הפעלת ה-DW-SORTeD בתצורות שונות וקצבי שידור שונים, כאשר מבצעים 2 איטרציות באלגוריתם ה-SORTeD. נשים לב, כי עבור שני מאורעות לבלוק האלגוריתם האופטימלי מתלכד עם התת-אופטימלי, כיוון שהמאורע האחרון הינו מקובע לקצה הבלוק (ר' סעיף 7.3.5). תוצאות ה-DW-SORTeD מושווים עם אלגוריתם ישיר של הסרת פרמטרי עירור של כל מסגרת שנייה (בקידוד) והשלמתם ע"י אינטרפולציה (במפענח). הביצועים של אלגוריתם ישיר זה מוצגים בשורה אחרונה של הטבלה. המדידות שנעשו לבחינת הביצועים הם – שורש של השגיאה הריבועית הממוצעת (RMS) ואחוזי המסגרות החריגות, כאשר ההשואה נעשית ביחס לערכים המקוריים שעברו קיצוץ לטווח הערכים המקונטים (clipping) בהתאם לתקן MELP-2400.

חריגות ה-pitch (outliers) מחושבות כאחוז המסגרות הקוליות מסה"כ המסגרות המקיימות $0.25 \le (F-\hat{F})/F < 0.5$. החריגות הגדולות (gross outliers) של ה-pitch מחושבות כאחוז

המסגרות הקוליות (מסך הכל מסגרות) המקימות: $(F-\hat{F})/F>0.5$, כאשר F תדר ה-pitch המקורי ו- $\hat{F}$ הוא תדר ה-pitch המקוונט (לאחר ה-TD). חריגות האנרגיה (outliers) מחושבות כאחוז המסגרות שעבורם: $5\le|G-\hat{G}|<10$ וגם $G>20$ או, לחילופין, $5\le\hat{G}-G<10$ ו- $G\le 20$. (המספרים הם ב-dB והם מותאמים לערכי האנרגיה המופקים ע״י מקודד MELP, שהטווח האפשרי שלהם הוא [10dB – 77dB].) כלומר, אם האנרגיה המקורית נמוכה מסף מסוים, הקטנתה הנוספת לא תיחשב כחריגה. החריגות הגדולות (gross outliers) של האנרגיה הם אחוז המסגרות המקיימות: $|G-\hat{G}|\ge 10$ וגם $G>20$ או, לחילופין, $\hat{G}-G\ge 10$ ו- $G\le 20$.

Table 7-a. Performance of excitation quantization by Temporal Decomposition.

טבלה 7-א. ביצועים של קוונוט פרמטרי העירור באמצעות ה-TD.

Joint TD	*M/N*	Phi Cdbk	Rate [bps]	Energy			Pitch		
				RMS, [dB]	Outliers, [%]	Gross Outliers, [%]	Avg Rel. Err. [%]	Outliers, [%]	Gross Outliers, [%]
Yes	7/3	5	355	3.36	7.17	1.00	2.01	0.25	0
Yes	7/3	4	336	3.52	9.17	1.10	2.08	0.25	0.04
Yes	7/3	2	298	4.13	14.87	2.19	3.1	0.61	0
Yes	11/4	6	331	3.20	7.88	0.68	2.3	0.21	0
Yes	11/4	5	315	3.37	9.17	0.86	2.4	0.21	0
Yes	11/4	3	283	3.82	13.44	1.22	3.1	0.54	0
Yes	11/3	6	263	4.17	15.30	2.69	3.3	0.29	0
No	7/3	2	342	3.25	6.24	0.90	6.06	2.90	0.68
No	11/5/3[3]	1	295	3.43	9.5	0.75	6.07	3.05	0.54
No	11/3	2	262	4.861	20.947	4.555	6.06	2.91	0.65
1:2 Decimation			222	4.92	16.99	5.48	7.21	5.77	2.54

מודל ה-DW-SORTeD מצליח לקבל אחוזי התאמה סבירים באיזור של 300 סיביות לשנייה (כולל סיבית מידע הצד (UV/V) עבור כל מסגרת ומסגרת). ירידה נוספת במס׳ הסיביות גורמת לדגרדציה מהירה, בייחוד באנרגיה, עקב אי-היכולת לחקות את ההשתנות המהירה ממסגרת למסגרת של האנרגיה. ה-pitch בגלל אופיו החלק מקבל דגרדציה קטנה יותר, לכן ניתן בסכימה של ה-TD אשר מופעל על ה-pitch והאנרגיה בנפרד, לאפשר יותר מאורעות עבור האנרגיה מאשר עבור ה-pitch. אמנם, הבעייתיות של מודל ה-TD הנפרד לכל פרמטר בכך, שהיא צורכת מספר כפול של סיביות לתיאור פונקציות המאורע, וכך מעלה את קצב השידור משמעותית. זה מחייב הקצאה

[3] 5 מאורעות עבור האנרגיה ו-3 מאורעות עבור ה-pitch.

מועטה במיוחד של הסיביות לקידוד צורת פונקצית המאורע. בפועל תצורת הקידוד המועדפת היא 4 מאורעות בבלוק של 11 סיביות או 3 מאורעות בבלוק של 7 סיביות (בהתאם להשהיה אלגוריתמית רצויה), כאשר הקידוד נעשה על ה-pitch והאנרגיה במשותף.

פרק 8

מקודד דיבור בקצבים נמוכים מאד המבוסס על MELP ו-DW-SORTeD

8.1 הקדמה

בפרקים הקודמים תיארנו איך ניתן להשתמש ב-TD לדחיסת המעטפת הספקטרלים ופרמטרי העירור. בפרק זה נציג מקודד דיבור בקצב קבוע של כ-600-700 bps, בהתבסס על שימוש באלגוריתם ה-SORTeD. המקודד מבצע אנליזה וסינתזה בצורה דומה למקודד העירור המעורב (MELP) בקצב 2400 bps וההבדל העיקרי בין שני המקודדים הוא בביצוע הקוונטיזציה וצמצום של חלק מפרמטרי העירור. המקודד המוצע משתמש באלגוריתם ה-DW-SORTeD שהוצג בפרקים הקודמים לשם ביצוע הקוונטיזציה. מוצעות שתי גרסאות של המקודד: עם השהיה אלגוריתמית של 11 מסגרות ושל 7 מסגרות. המקודד נבדק ע"י מדד ה-PESQ, להערכת איכותו הסובייקטיבית.

8.2 מקודד מבוסס MELP ל-600 bps

8.2.1 כללי

סכימה מלבנית של המקודד המוצע ניתנת באיור 8-א.

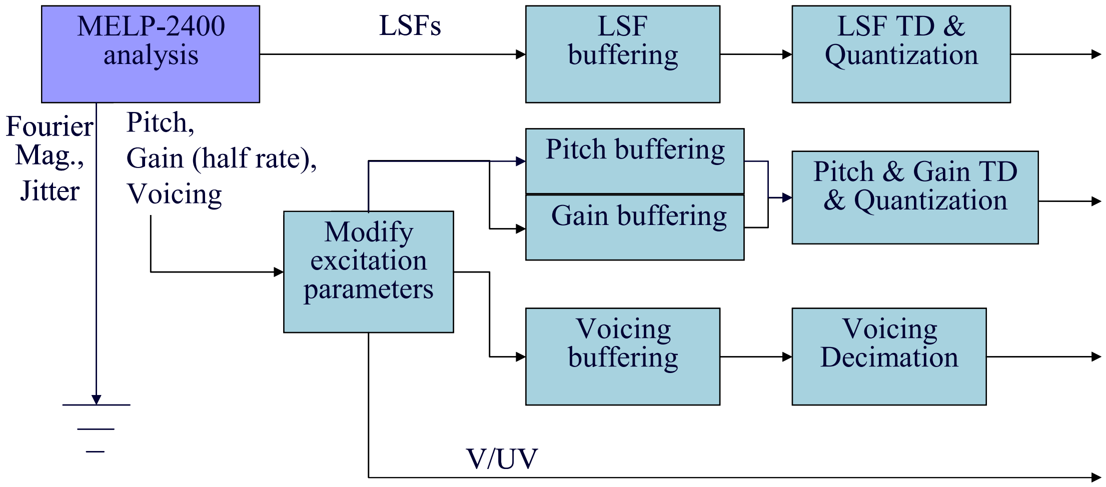


Figure 8-1. Block diagram of 600 bps vocoder

איור 8-א. סכימה מלבנית של המקודד 600 bps

לאחר אנליזת ה-MELP (סעיף 2.5) , חלק מפרמטרי העירור, כגון jitter והספקטרום פולס העירור אינם משודרים כלל. חלק אחר (כמו קוליות של פסי תדר שונים) מקוונטים בצורה גסה, כפי שיוצג בסעיף הבא. המעטפת הספקטרלית ופרמטרי העירור העיקריים (pitch ואנרגיה) מקוונטים בעזרת אלגוריתם ה-DW-SORTeD, אשר הוצג בפרק 5 (לשימוש עבור המעטפת הספקטרלית) ובפרק 7 (לשימוש עבור פרמטרי העירור). ראינו בפרק 7, כי ה-pitch מקודד ללא אינפורמציה אם מסגרת כלשהי היא קולית או א-קולית. לכן, יש לשדר בנוסף ל-pitch סיבית שמסווגת כל מסגרת להיות קולית או א-קולית (ניזכר כי סיביות ה-voicing הנוספות מתארות את "מידת הקוליות" עבור מסגרות המסווגות כקוליות). לפני הכניסה לשלב של TD וקוונטט פרמטרי העירור (pitch ואנרגיה) עוברים מספר שינויים (ר' סעיף 7.3.2) : קודם כל, משתמשים רק במדידת אנרגיה אחת לכל מסגרת דיבור, במקום שניים, כמוגדר בתקן. בנוסף לכך, החלטות קולי א-קולי משתנות, כך שיותר מסגרות יסווגו כא-קוליות. באזורים המסווגים כא-קוליים, נקבע ערך ה-pitch ע"י אינטרפולציה לינארית של המסגרות הקוליות הסובבות אותם, ע"מ לקבל מסלול ההשתנות של ה-pitch רציף.

8.2.2 קוונטיזציה של תבנית ה-voicing (קוליות).

בתקן ה-MELP לקידוד אות דיבור בקצב 2400 bps [6], משתמשים בארבע סיביות כדי לציין את סוג העירור בארבעת פסי התדר העליונים וסיבית נוספת מקודדת במשותף עם ה-Pitch ומציינת האם המסגרת היא קולית או א-קולית. כאשר מחליטים שתחום התדר הנמוך הוא א-קולי, מאלצים גם את כל תחומי התדר האחרים להיות א-קוליים וכאשר מחליטים שתחום התדר הנמוך קולי ניתן לקבל החלטת קוליות שונה בכל אחד מארבעת תחומי התדר הנותרים. McCree וחבריו הציעו [61] להשתמש בשתי סיביות במקום ארבע, כלומר לבצע קוונטיזציה לתבנית ה-voicing. בבדיקה

סטטיסטית שנעשתה למספר הופעות תבנית voicing מסוימת בקטעי דיבור רועשים ושקטים נמצא כי ישנן תבניות voicing שמופיעות בהסתברות גבוהה יותר מתבניות אחרות. בהתאם לכך נבחרו תבניות ה-voicing הבאות: 1000, 1100, 1110, 1111, כאשר המילה $b_2b_3b_4b_5$ מתארת את החלטת הקוליות בתחומי התדר השונים. לדוגמא, 1000 מתאר את התבנית הבאה: פס התדר 0-500Hz (b_1) במצב קולי (כמו בכל המסגרות הקוליות), פס התדר 500-1000Hz (b_2) במצב קולי, ושאר פסי התדר: 1000-2000Hz, 2000-3000Hz ו-3000-4000Hz במצב א-קולי. תבניות אילו מתארות כולן את המצב בו התדרים הנמוכים הם קוליים, והתדרים הגבוהים הם א-קוליים, כלומר ישנו תדר קיטעון אחד, המפריד בין המרכיב הקולי בתדרים הנמוכים לבין מרכיב הרעש בתדרים הגבוהים. ייצוג כזה נפוץ במודלים אחרים של אות העירור, כגון מודלים סינוסואידליים [62,4]. תדר זה הוא פרמטר אשר משתנה לאט יחסית, וחשוב מבחינת השמיעה האנושית, שלא ישתנה בצורה חדה ממסגרת למסגרת [62]. שימוש באינטרפולציה של תדר הקיטעון (שמקבל ערכים של 1000, 2000, 3000, 4000 Hz), ניתן להוריד את מספר הסיביות לייצוג הקוליות ע"י דילוג על מסגרות מסוימות והשלמת המידע ע"י האינטרפולציה של תדר הקיטעון בצד המקלט. אנו מצאנו, כי ניתן לצמצם כך פי שלושה ויותר את כמות הסיביות המושקעת בקידוד תבנית הקוליות ללא פגיעה משמעותית באיכות הדיבור עקב דילוגים אלו. במקודד הנוכחי הוחלט לקודד את תבנית הקוליות של שלוש מסגרות בכל בלוק של 11 מסגרות או 2 מסגרות בבלוק של 7 מסגרות. הקוונטיזציה נעשית באופן הבא:

Table 8-a. Quantization table of voicing parameters.

טבלה 8-א. טבלת קוונטיזציה של תבנית ה-voicing.

תדר קיטעון [Hz]	מילת קוד	תבנית voicing מקודדת $b_2b_3b_4b_5$	תבנית voicing מקורית $b_2b_3b_4b_5$
1000	00	1000	0000
1000	00	1000	0001
1000	00	1000	0010
1000	00	1000	0011
1000	00	1000	1001
1000	00	1000	1010
1000	00	1000	1000
2000	01	1100	0100
2000	01	1100	1101
2000	01	1100	1100
2000	01	1100	0101
3000	10	1110	0110
3000	10	1110	1110
4000	11	1111	0111
4000	11	1111	1011
4000	11	1111	1111

מסגרות, שעבורם משודרת תבנית הקוליות נבחרות להיות מפוזרות בצורה אחידה על פני החלק הקולי של בלוק האנליזה. במסגרות הקוליות, עבורם לא משודרת מידע על תבנית הקוליות, נקבע תדר קיטעון ע"י אינטרפולציה לינארית של תדר הקיטעון בין המסגרות הקוליות השכנות בעלות תבנית קוליות ידועה. לאחר מכן הוא מקוונט ל-2 סיביות (מקבל אחד מהערכים האפשריים של 1000, 2000, 3000, 4000 Hz, אשר מתורגמים לאחת התבניות בטבלה 8-א).

8.2.3 הקצאת הסיביות למקודד 600-650 bps

מספר סכימאות של הקצאות הסיביות נבחנו עבור המקודד שתואר לעיל, אשר מביאות לקצב כולל שנע בין 600 ל-680 סיביות לשנייה וגודל בלוק אנליזה של 7 ו-11 מסגרות (השהיה האלגוריתמית). אנו נביא כאן לדוגמה את אחת הסכימאות, עם בלוק אנליזה של 7 מסגרות ואחת ההקצאות עם בלוק אנליזה של 11 מסגרות.

הקצאת הסיביות של המקודד עם גודל הבלוק של 11 מסגרות ניתנת בטבלה 8-ב בהינתן כי קצב המסגרות הנו 44.44 Hz, מקבלים קצב כולל של 602 סיביות לשנייה. בסכימה זו וקטורי המטרה של הפרמטרים הספקטרליים מקודדים ב-18 סיביות (10 סיביות עבור ארבעת ה-LSF הנמוכים ו-8 סיביות עבור השישה הנותרים). פונקציות המאורע מקודדים ע"י 5 סיביות כל אחת (שתיים עבור קידוד הצורה באמצעות VQ, ועוד שלושה עבור אורכיהן). יש לקודד 3 מאורעות לבלוק של הפרמטרים הספקטרליים.

פרמטרי העירור מקודדים בצורה הבאה: 3 מסגרות מתוך הבלוק מכילות תבנית הקוליות (שתי סיביות כל אחת), אנרגיה ו-pitch עוברים קוונטיזציה באמצעות אלגוריתם ה-DW-SORTeD, במשותף, כאשר וקטורי המטרה מקוונטים ע"י קוונטיזציה סקלרית עם 5 סיביות לכל אחד מהפרמטרים וישנם 4 מאורעות בתוך כל בלוק באורך 11 מסגרות. צורת כל פונקצית מאורע מקוונטת ע"י 4 סיביות (VQ). ע"מ לפשט את האלגוריתם ולהקטין את כמות הסיביות, נקבע, כי המאורע האחרון לא משנה את מקומו בשלב החיפוש אחר הסגמנטציה האופטימלית, והוא ממוקם תמיד במסגרת אחרונה של הבלוק. תנאי זה מביא, למעשה, לכך שהחפיפה בין בלוקים עוקבים תהיה תמיד שווה למסגרת אחת, ואין צורך לשדר מיקום של המאורע האחרון (או, לחילופין, אורך הסגמנט האחרון). למעשה יש בדיוק $\binom{10}{3}=120$ אפשרויות למקם את 3 המאורעות הנותרים בתוך הבלוק. כלומר, ניתן להעביר את כל אורכי הסגמנטים ב-7 סיביות. מידע צד נוסף, שיש לשדר לפענוח פרמטרי העירור, הוא החלטות קולי/א-קולי, אשר צורכות סיבית אחת לכל מסגרת.

Table 8-b. . Bit allocation for MELP-based speech coder. The allocation is given for the joint coding of 11 frames.

טבלה 8-ב. הקצאת סיביות למקודד מבוסס MELP. ההקצאה היא עבור קידוד של 11 מסגרות.

Param.	Bits/Block (11 frames)	Bit- Rate [bps]
LSF (3 events)	(10+8+2+3)*3=69	278.8
Gain & Pitch (4 events)	(5+5+4)*4+7= 63	254.6
V/UV	11	44.4
Voicing(3 frames)	2*3	24.2
Total	149	602

הקצאת סיביות של מקודד אופייני עם גודל בלוק של 7 מסגרות ניתנת בטבלה 8-ג. בהינתן כי קצב המסגרות הנו 44.44 Hz, מקבלים קצב כולל של 634.8 סיביות לשנייה. כאן משתמשים ב-2 מאורעות

לספקטרום ו-3 מאורעות לעירור (קוונטיזציה באמצעות DW-SORTeD). וקטורי המטרה ופונקציות המאורע של הספקטרום מקבלים אותה הקצאה, כמו בסכימה הקודמת. רק 2 מסגרות עם תבניות הקוליות משודרות. וקטורי מטרה של אות העירור מקוונטים כמו בסכימה הקודמת (קוונוט סקלרי של 5 סיביות לכל אחד מהרכיבים – האנרגיה וה-pitch), אך צורת פונקציות המאורע מקוונטת ב-3 סיביות בלבד. גם כאן מקבעים את מיקום המאורע האחרון, ולכן יש סה"כ $\binom{6}{2}=15$ אפשרויות למקם מאורעות בתוך הבלוק, כלומר מספיקות 4 סיביות לעשות זאת.

Table 8-c. . Bit allocation for MELP based speech coder. The allocation is given for the joint coding of 7 frames.

טבלה 8-ג. הקצאת סיביות למקודד מבוסס MELP. ההקצאה היא עבור קידוד של 7 מסגרות במשותף.

Param.	Bits/Block (7 frames)	Bit Rate [bps]
LSF(2 events)	(10+8+2+3)*2=46	292
Gain & Pitch (3 events)	(5+5+3)*3+4= 43	273
V/UV	7	44.4
Voicing(2 frames)	2*2	25.4
Total	100	634.8

8.2.4 הקצאת הסיביות למקודד 800-880 bps

לצורך השוואת הביצועים עם המקודד הסופי ובחינת הדגרדציה, שמכניסה קוונטיזציה עמוקה של פרמטרי העירור באמצעות ה-DW-SORTeD, נגדיר משפחת מקודדי דיבור מבוססי MELP בעלי קצבים בתחום 800-880 bps. במקודדים אלה המעטפת הספקטרלית מקוונטת באמצעות ה--DW SORTeD (הקצאות הסיביות בהתאם לסעיף 8.2.4). פרמטרי העירור מכוונים גם הם בהתאם ל-8.2.4, אך אין מבצעים קוונטיזציה של pitch ואנרגיה באמצעות ה-DW-SORTeD, אלא מקוונטים אותם קוונטיזציה סקלרית (5 סיביות יוקצה לכ"א מפרמטרים אלה). מקודדים אלה יסייעו להעריך עד כמה שימוש ב-TD לצורך קוונטיזציית אות העירור משפיע על האיכות הכוללת של הדיבור.

8.3 בחינת ביצועים

8.3.1 סכימאות קידוד להשוואה

הביצועים של המקודד הוערכו באמצעות מדד ה-PESQ עם מיצוע על פני 10 משפטים מאוזנים פונטית מתוך מאגר ה-TIMIT , והושוו ל-4 מערכות ייחוס :

- מקודד ה-MELP התקני (2400 סיביות לשנייה)
- מקודד בעל עירור ה-MELP המצומצם (בהתאם לסעיף 8.2.4) וקידוד המעטפת התקני (1111 סיביות לשנייה עבור המעטפת). מקודד זה הנו בעל קצב כולל של 1624 סיביות לשנייה ויכונה בהמשך MELP-1600.
- מקודד בעל עירור ה-MELP המצומצם ודחיסת המעטפת באמצעות DW-SORTeD (סעיף 8.2.4). נבחנו אורכי הבלוק של 7 ו-11 מסגרות, המכונים Sp-DW-SORTeD-7 ו-Sp-DW-SORTeD-11 בהתאמה. קצבים אופייניים של סכימה זו הם 800-880 סיביות לשנייה.
- מקודד דיבור מבוסס MELP, אשר מבצע קוונטיזציה במשותף של הפרמטרים עבור 4 מסגרות עוקבות במשותף, בהתאם ל-[57] עם ההבדל היחיד של אורך מסגרת האנליזה. במקודד המוצע ב-[57] בוחרים להשתמש באורך המסגרת של 25 מילישניות, ובמקודד ששימש להשוואה השתמשו באורך המסגרת המקורית של MELP-2400 (22.5 מילישניות). המקודד המושווה הנו בעל קצב קבוע של 666 סיביות לשנייה ויכונה MELP-666. מקודד זה מומש במסגרת פרויקט סטודנטים במעבדה לעיבוד אותות ותמונות בטכניון.

8.3.2 תוצאות ההשוואה

התוצאות עבור מספר סכימאות קידוד ניתנות בטבלה 8-ד ובאיור 8-ב. בטבלה 8-ד מוצגים ביצועים של שני המקודדים שפורטו בהרחבה בסעיף 8.2.3 בהשוואה למערכות הייחוס המתוארות לעיל. מקודדים אלו יכונו DW-SORTeD-7 ו-DW-SORTeD-11 עבור גודל בלוק האנליזה של 7 ו-11 מסגרות בהתאמה.

Table 8-d. Performance estimation of VLBR coders. The standard deviation is about 0.2

טבלה 8-ד. הערכת ביצועים של מקודדי דיבור בקצבים נמוכים מאד. סטיית התקן של כ"א מהמדידות היא כ-0.2

Coding scheme	Bit Rate [bps]	Average PESQ score
DW-SORTeD-7	635	2.55
Sp-DW-SORTeD-7	825	2.64
DW-SORTeD-11	602	2.61
Sp-DW-SORTeD-11	812	2.69
MELP-2400	2356	3.13
MELP-1600	1624	2.95
MELP-666	666	2.33

המקודדים המוצגים בטבלה לעיל יחד עם מקודדים נוספים (סכימאות הקצאת סיביות נוספות) מוצגים באיור 8-ב. בחלק ממקודדי ה-DW-SORTeD (המסומנים DW-SORTeD-7 ו-DW-SORTeD-11 בגרף), פעולת ה-TD של פרמטרי העירור (pitch ואנרגיה) נעשתה בנפרד (מקודדים אלה מסומנים באות "s" באיור 8-ב). הדבר תורם ליציבות של המערכת, אך מגדיל בצורה משמעותית את כמות הסיביות שיש להשקיע בקידוד העירור, כיוון שיש לשדר פעמיים מידע על צורה ואורך של פונקציות המאורע. ע"מ להגיע לקצבים הרצויים במקרה זה, יש לצמצם את כמות הסיביות שיש להשקיע ע"י ייצוג מינימליסטי של צורות פונקציות המאורע ולהקטין את מספר המאורעות עבור פרמטרי העירור. ניתן, לחילופין או במקביל, להקטין את הקצאת הסיביות עבור המעטפת הספקטרלית אך הדבר עלול לפגוע במובנות של הדיבור המסונטז.

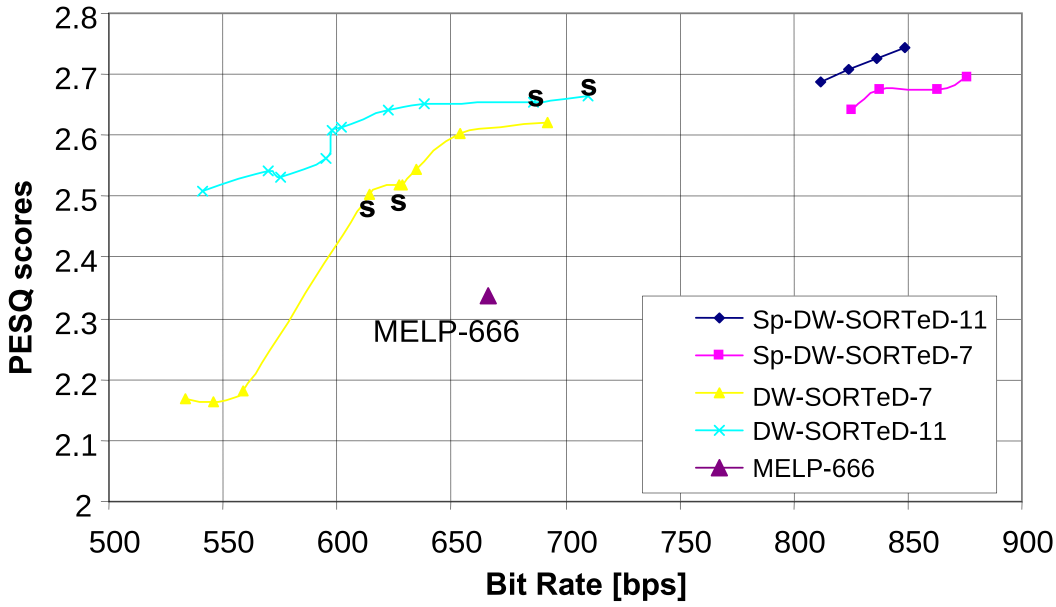


Figure 8-2. Performance estimation of VLBR coders. "S" denotes separate pitch and energy quantization for DW-SORTeD.

איור 8-ב, הערכת ביצועים של מקודדי דיבור בקצבים נמוכים מאד. "S" מציין מערכות בהם קידוד ה-pitch והאנרגיה נעשה בנפרד עבור מקודדי ה-DW-SORTeD.

8.3.3 דיון ומסקנות

השוואה עם האלגוריתם הישיר של MELP-666 (סעיף 8.3.1), בעל קצב שידור הקרוב לקצבים של המקודדים מבוססי DW-SORTeD, מעידה על שיפור ניכר. האלגוריתמים DW-SORTeD-7 ו-DW-SORTeD-11 טובים ממנו ב-0.2 ו-0.3, בהתאמה, בקצבי שידור הקטנים מ-666 סיביות לשנייה. השיפור בא לידי ביטוי גם במובנות וגם באיכות הצליל.

נבחן כעת באיזו מידה דחיסה עמוקה של פרמטרי העירור והמעטפת הספקטרלית וכן צמצום של פרמטרי עירור אחדים פוגעים באיכות הנתפסת של האות המסונטז (ר' איור 8-ג).

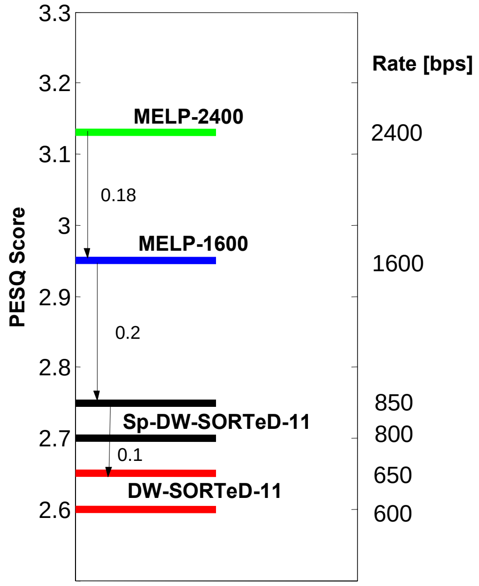


Figure 8-3. Perceptual quality degradation for different VLBR algorithms

איור 8-ג. פגיעה באיכות הדיבור כתוצאה מהורדת קצב הסיביות עבור אלגוריתמי קידוד שונים.

מהאיור רואים כי השמטה של פרמטרי MELP מסוימים, כגון אמפליטודות פורייה ו-jitter, יחד עם דחיסת תבנית ה-voicing, מביאות לדגרדציה של 0.18 במדד ה-PESQ. דגרדציה זו באה לידי ביטוי בעיקר בעיוותים, כגון הגברת זמזומים (מתכתיות) של המוצא. דחיסה של המעטפת (280-315 סיביות לשנייה במקום 1111 סיביות לשנייה) מביאה לדגרדציה נוספת של 0.2-0.25. דגרדציה זו מודגשת בעיקר בצליליים אנפיים (כגון m, n). תוצאה זו עקבית עם הבדיקות שתוארו בסעיף 6.3.3. לבסוף, הפעלת ה-DW-SORTeD על ה-pitch והאנרגיה מביאה לדגרדציה נוספת של כ-0.1. דגרדציה זו מתבטאת בשינויים קלים של עקום ה-pitch (לעיתים לא מורגש ע״י אוזן לא מאומנת) ושינויי עוצמה נקודתיים.

לסיכום, דחיסת המעטפת הספקטרלית פוגעת במידה פחותה יותר באיכות הנתפסת, מאשר צמצום ודחיסת פרמטרי העירור (מסקנה זו אומתה גם ע״י השמיעה לא פורמלית של האותות המסונטזים). הבעייתיות של דחיסת פרמטרי העירור של MELP במלואם נובעת מריבוי פרמטרים שהם בעלי תלויות זמניות שונות. נשים לב, כי צמצום פרמטרים אלו הביא לדגרדציה גדולה יותר מאשר דחיסת ה-pitch ואנרגיה באמצעות ה-TD. ניתן לומר, שאלגוריתם ה-SORTeD הצליח לדחוס את פרמטרי העירור העיקריים עם דגרדציה קטנה.

נשים לב, בנוסף, שהשימוש באורך בלוק של 11 מסגרות משפר את הביצועים בהשוואה לבלוק של 7 מסגרות. ככל שמקטינים את גודל בלוק האנליזה ומספר המאורעות בבלוק, מאבדים מעצמתה של הפעולה של ה-TD, הבנויה במקור עבור חוצצים ארוכים יותר.

פרק 9

סיכום והצעות להמשך המחקר

9.1 סיכום

עבודה זו מתמקדת בקידוד דיבור בקצבים נמוכים מאד, על בסיס חיזוי לינארי (LPC), כאשר הוצבה המטרה לפתח מקודד דיבור בקצב קבוע של כ-600 סיביות לשנייה ו/או קידוד המעטפת הספקטרלית בקצב של כ-300 סיביות לשנייה.

לצורך השגת היעד, פותחה טכניקה חדשה לביצוע הפירוק של רצף וקטורי הכניסה (שהם יכולים להיות וקטורי המעטפת הספקטרלית, המיוצגים ע״י LSF או פרמטרי עירור) המבוססת על מודל ה-Temporal Decomposition (TD) לייצוג אותות הדיבור. טכניקה זו, המכונה DW-SORTeD (Dynamically Weighted Sub-Optimal Restricted Temporal Decomposition), מאפשרת בצורה יעילה, מבחינה חישובית, לייצג בלוק של וקטורי כניסה ע״י מספר מצומצם של נציגים (וקטורי המטרה) ומספר תואם של פונקציות אינטרפולציה, כאשר הקריטריון למינימיזציה הוא סכום השגיאות הריבועיות המשוקללות עם משקלים דינמיים מתאימים (משקלים התלויים בוקטורי הכניסה). אלגוריתם זה פותח בהתבסס באלגוריתם ה-ORTD הנועד לדחיסת ספקטרום דיבור שלא בזמן אמת, עקב סיבוכיותו הגבוהה. במאמץ לפיתוח האלגוריתם המשופר, ה-ORTD הותאם לקריטריון הסכום המינימלי של השגיאות הריבועיות המשוקללות, עם משקולות משתנות בזמן ונבחרו משקלות המתאימים לדחיסה יעילה של פרמטרי ה-LSF היעילה. בהמשך, פותחה הגרסה התת-אופטימלית, שהביאה להורדה משמעותית בסיבוכיות האלגוריתם, דבר המאפשר את יישומו במערכות זמן אמת, תוך פגיעה מזערית באיכות ההתאמה של המעטפת הספקטרלית.

יישום של טכניקה זו על וקטורי ה-LSF הניב מערכת לקידוד המעטפת הספקטרלית (המיוצגת ע״י וקטורי ה-LSF) ב-6.75-8.3 סיביות למסגרת דיבור באורך 22.5 מילישניות. במערכת קידוד ה-LPC שיטה זו מקודדת את המעטפת הספקטרלית בקצבים של 300-370 סיביות לשנייה באיכות התאמה של 2.1-2.25 dB בממוצע (LSD). אלו הם הערכים האופייניים של מקודדי המעטפת הספקטרלית הפועלים בקצבים של 500-600 סיביות לשנייה. תכונת החלקות של הספקטרום המתקבל, המובנית בתוך האלגוריתם, מביאה לדגרדציה נמוכה של אות הדיבור במוצא המערכת. שילוב של קידוד DW-SORTeD עם ייצוג וקוונוט העירור של מקודד ה-MELP-2400 התקני מפיק דיבור בקצב 1.6 Kb/s באיכות של 2.8 (PESQ), כאשר קידוד מלא (2.4 Kb/s) של MELP הפיק אותות מוצא באיכות ממוצעת של 3.0 (PESQ).

בהמשך למאמץ לפיתוח המקודד בקצב של כ-600 סיביות לשנייה, האלגוריתם DW-SORTeD הותאם לשימוש לדחיסה עמוקה של פרמטרי עירור, כגון pitch ואנרגיה. ביצוע של פעולת הנרמול בכניסה למערכת מאפשר ביצוע קוונטיזציה של ה-pitch והאנרגיה במשותף, בקצבים של 260-320 סיביות לשנייה, עבור קצב עדכון פרמטרים של 44.44 Hz.

על סמך הרעיונות שפותחו בעבודה זו ובהתאם לתוצאות שהתקבלו בחלקים הקודמים מוצע מקודד דיבור המבוסס על אנליזת MELP, עם נקודות עבודה בין 600 ל-650 סיביות לשנייה, אשר מיועד לערוץ תקשורת חד-כיווני (Half-Duplex) בגלל השהייתו הגבוהה יחסית (השהיה אלגוריתמית של המקודד היא 160 – 250 מילישניות). ביצועי המקודד המוצע נבדקו לעומת מקודד בקצב דומה המבוסס על קוונטיזציה מטריצית (MQ) של 4 מסגרות עוקבות של תקן MELP (שהוצע באחרונה בספרות) ונמצא כי ציון ה-PESQ של האלגוריתם החדש גבוה ב-0.2-0.3.

שימוש בטכניקת ה-TD עם קריטריון השגיאה המותאם לאוזן האנושית מאפשר הורדת קצב הסיביות לקידוד המעטפת הספקטרלית, מבלי להביא לדגרדציה משמעותית בהתאמת הספקטרום (ציון ה-LSD הממוצע). יש לציין, כי למרות האחוז הגבוה יחסית של המסגרות החריגות (מדדי LSD מעל 4dB ומעל 2dB), אין אות המוצא מתאפיין בעוותי ספקטרום מקומיים רבים, בגלל תכונת ההחלקה המובנית של אלגוריתם ה-DW-SORTeD. להבדיל משיטות קידוד מעטפת דיבור חסרות זיכרון, בהן חריגים כאלה מהווים בד"כ הפרעות אקראיות, הנובעות מרעש קוונטיזציה, ונתפסות ע"י האוזן האנושית, בשיטת ה-TD הציונים החריגים יכולים להיווצר עקב החלקת הספקטרום על פני הזמן ולכן לא בהכרח מתורגמים להפרעות שנתפסות ע"י האוזן. הדבר מסביר את הציונים הגבוהים יחסית של הערכת MOS, למרות כמות ניכרת של המסגרות החריגות. שימוש ב-TD לדחיסת הדיבור מאפשר להוריד את קצב העדכונים של המידע הספקטרלי קרוב לקצב הפונמות בדיבור האנושי (כ-12 לשנייה), כלומר מתקרב לקצה היכולת של מקודדי דיבור פרמטריים המשמרים את זהות הדובר (למעט מקודדי דיבור פונטיים, הנועדים לדובר מסוים), שכן לא ניתן להוריד את כמות העדכונים הספקטרליים מתחת למספר מאורעות הדיבור האמיתיים (הפונמות).

שימוש מוצלח ב-DW-SORTeD בדחיסת פרמטרים שהם אינם פרמטרים ספקטרליים והתלות ביניהם פחות ברורה (האנרגיה וה-pitch), הדגים את היכולות הנוספות של אלגוריתם ה--DW-SORTeD, אשר יכול לשמש גם לצורך מידול ודחיסה של וקטורי כניסה כלליים, תוך ניצול האינווריאנטיות להתמרה אפינית של וקטורי כניסה ואפשרות של שימוש בשגיאה משוקללת המשתנה בזמן.

הסיבוכיות של האלגוריתם ה-DW-SORTeD , המותאמת ליישומי זמן אמת, מאפשרת שימושן מקודדי דיבור מעשיים וכן ביישומים נוספים.

9.2 הצעות להמשך מחקר

קיימים מספר כוונים להמשך המחקר בנושא קידוד הדיבור בעזרת מודל ה-TD.

מידת הדגרדציה שנגרמת עקב דחיסה והשמטה של חלק מפרמטרי העירור עולה על זו הנגרמת עקב דחיסת המעטפת הספקטרלית באותו היחס. לכן, יש לבחון מודלי עירור נוספים ושילובם עם גישת ה-TD (כגון המודל הסינוסואידלי [62] או מודל ה-Multi-Band Excitation [3]) לשיפור איכות העירור המתקבל בכ-300 סיביות לשנייה. מודלים אלה יכולים להתאים יותר לדחיסה עמוקה, עקב כמות קטנה יותר של פרמטרים מסוגים שונים במודל העירור וההשתנות החלקה של חלק מהם (כגון תדר הקיטעון ב-[62]).

העבודה הנוכחית התמקדה בנקודת עבודה של קצבים נמוכים מאד, אך ניתן לנצל את תכונות ה-TD גם במקודדי דיבור בעלי קצב סיביות ואיכות גבוהים יותר. במקודדים אלה שימוש ב-TD עשוי לאפשר הגדלת הרזולוציה הזמנית של הפרמטרים הספקטרליים וכך להעלות את איכות הדיבור המופק מבלי להגדיל את קצב השידור.

התרה של קצב שידור משתנה במקודדים מבוססי DW-SORTeD תוכל אף להביא, בסבירות גבוהה, למקודדים בקצב ממוצע נמוך מזה שהושג בעבודה הנוכחית וכן לאיכות משופרת. חלק ניכר ממאמץ מחקרי כזה הוא פיתוח קריטריונים (מבוססי SFTR או קריטריונים אחרים) להערכת מספר המאורעות בבלוק האנליזה וקביעת מיקומם ההתחלתי.

סוגיה חשובה נוספת קשורה במחקר המכוון להפחתת ההשהיה המובנית באלגוריתם DW-SORTeD תוך שמירה על קצבי שידור נמוכים ככל הניתן. זאת משום שהביצועים של האלגוריתם שפותח בעבודה הנוכחית יורדים עם הקטנת אורך הבלוק/השהיה אלגוריתמית של המקודד. יחד עם זאת ההשהיה של המערכת היא גבוהה משמעותית מזו המקובלת כיום במקודדים המסחריים.

נספח א'

עידון וקטורי המטרה האופטימלי במובן של שגיאה ריבועית משוקללת מינימלית עם משקלים דינמיים

בשלב של עידון של וקטורי המטרה יש לעדן את וקטורי המאורע ע"י מינימיזציה של שגיאת הבלוק הכוללת כפונקציה של וקטורי המטרה, בהינתן פונקציות המאורעות. למעשה, יש למצוא מטריצת וקטורי המטרה הממזערת את מדד ה-WMSE לבלוק, בהינתן פונקציות המאורעות $\phi_{i,n} \triangleq \phi_i(n)$ ומיקומי המאורעות n_i, ז"א יש למזער את:

(i)

$$E(\mathbf{A}) = \sum_{k=0}^{k=M-1} \sum_{n=n_k}^{n_{k+1}-1} g(\mathbf{a}_k, \mathbf{a}_{k+1}, n),$$

כאשר

(ii)

$$g(\mathbf{a}_k, \mathbf{a}_{k+1}, n) = \sum_{k=0}^{k=M-1} \sum_{n=n_k}^{n_{k+1}-1} (\mathbf{y}(n) - \mathbf{a}_k\phi_k(n) - \mathbf{a}_{k+1}\phi_{k+1}(n))^T \mathbf{W}(n)(\mathbf{y}(n) - \mathbf{a}_k\phi_k(n) - \mathbf{a}_{k+1}\phi_{k+1}(n)),$$

כלומר יש לפתור את:

(iii)

$$\frac{dE(\mathbf{A})}{d\mathbf{A}} = \sum_{k=0}^{k=M-1} \sum_{n=n_k}^{n_{k+1}-1} \frac{\partial g(\mathbf{a}_k, \mathbf{a}_{k+1}, n)}{\partial \mathbf{a}_k} + \frac{\partial g(\mathbf{a}_k, \mathbf{a}_{k+1}, n)}{\partial \mathbf{a}_{k+1}} = 0.$$

הגזירה של הביטוי שבסכום נותנת:

$$\text{(iv)} \frac{\partial g(\mathbf{a}_k, \mathbf{a}_{k+1}, n)}{\partial \mathbf{a}_k} = 2\phi_k^2(n)\mathbf{W}(n)\mathbf{a}_k - 2\phi_i(n)\mathbf{W}(n)\left(\mathbf{y}(n) - \mathbf{a}_{k+1}\phi_{k+1}(n)\right), \quad 0 \le k \le M-1$$

ההצבה לתוך (iii) מניבה:

(v)

$$\left[\sum_{n=n_{k-1}}^{n_{k+1}-1}\phi_k^2(n)\mathbf{W}(n)\right]\mathbf{a}_k+\left[\sum_{n=n_{k-1}}^{n_k-1}\phi_k(n)\phi_{k-1}(n)\mathbf{W}(n)\right]\mathbf{a}_{k-1}+\left[\sum_{n=n_k}^{n_{k+1}-1}\phi_k(n)\phi_{k+1}(n)\mathbf{W}(n)\right]\mathbf{a}_{k+1}=$$

$$=\sum_{n=n_{k-1}}^{n_{k+1}-1}\phi_k(n)\mathbf{W}(n)\mathbf{y}(n)$$

כאשר $0 \le k \le M-1$.

את המשוואות הוקטוריות (v) ניתן לרשום בנפרד עבור כל אחד מ-p הרכיבים של וקטורי הפרמטרים הספקטרליים בנפרד (p הוא האורך של וקטורי הכניסה $(\mathbf{y}(n)$). מקבלים p מערכות משוואות לינאריות סימטריות ותלת-אלכסוניות:

$$\begin{pmatrix} d_{i,0} & x_{i,0} & 0 & \mathbf{0} \\ x_{i,0} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,M-1} & x_{i,M-1} \\ \mathbf{0} & 0 & x_{i,M-1} & d_{i,M} \end{pmatrix}\begin{pmatrix} a_{i,0} \\ \vdots \\ a_{i,M-1} \\ a_{i,M} \end{pmatrix}=\begin{pmatrix} b_{i,0} \\ \vdots \\ b_{i,M-1} \\ b_{i,M} \end{pmatrix},$$

$$\text{(vi)}\ d_{i,k}=\sum_n \phi_k^2(n)w_i(n),\quad x_{i,k}=\sum_n \phi_k(n)\phi_{k+1}(n)w_i(n),\quad b_{i,k}=\sum_n \phi_k(n)y_i(n)w_i(n),$$

$$1\le n\le N,\qquad 1\le i\le p$$

כאשר $a_{i,k}$ הינו הרכיב ה-i-י של וקטור המטרה ה-k-י. ב-(vi) הושמטו הגבולות מהסכומים, כיוון שהסכימה יכולה להיעשות על פני כל ציר הזמן, כיוון שהמכפילות בתוך הסכימה יתאפסו מחוץ לסגמנט הרלוונטי לפונקציות המאורע המשתתפות במכפילה.

אם הבלוקים עליהם מתבצע פירוק ה-TD הנם חופפים, אין לשנות מאורע האפס, כלומר אין לגזור לפי $\mathbf{a}_0$. במקרה זה המשוואה הראשונה בתוך (vi) נשמטת ו- (vi) הופך ל:

$$\begin{pmatrix} d_{i,1} & x_{i,1} & 0 & \mathbf{0} \\ x_{i,1} & \ddots & \ddots & 0 \\ 0 & \ddots & d_{i,M-1} & x_{i,M-1} \\ \mathbf{0} & 0 & x_{i,M-1} & d_{i,M} \end{pmatrix}\begin{pmatrix} a_{i,1} \\ \vdots \\ a_{i,M-1} \\ a_{i,M} \end{pmatrix}=\begin{pmatrix} b_{i,1}-x_{i,0}a_{i,0} \\ \vdots \\ b_{i,M-1} \\ b_{i,M} \end{pmatrix},$$

$$\text{(vii)}\ d_{i,k}=\sum_n \phi_k^2(n)w_i(n),\quad x_{i,k}=\sum_n \phi_k(n)\phi_{k+1}(n)w_i(n),\quad b_{i,k}=\sum_n \phi_k(n)y_i(n)w_i(n),$$

$$1\le n\le N,\qquad 1\le i\le p$$

נספח ב'

חישוב של מספר פעולות קריטיות בשלב של מציאת הסגמנטציה באלגוריתמי ORTD ו-SORTeD.

I. מספר השוואות

i. ORTD:

את מספר ההשוואות ניתן להפיק מדיאגרמת ה-trellis שמשמשת למציאת הסגמנטציה האופטימלית ב-ORTD (ר' איור 4-י). מספר ההשוואות בכל צומת שווה למספר הקשתות הנכנסות בו פחות אחד. בהינתן שבכל שלב של הדיאגרמה ישנם N-M+1 צמתים, נחשב את מספר ההשוואות שיש לעשות ע"מ לבחור מסלול מיטבי בכל צומת:

$$C_1 = (M-1)\sum_{i=1}^{N-M+1}(i-1) = (M-1)\frac{(N-M+1)(N-M)}{2}. \qquad \text{(i)}$$

לדוגמא, עבור $N=11,\ M=3$ מקבלים 72 השוואות.

ii. SORTeD:

נוכל לחשב רק מספר ממוצע של השוואות, כיוון שאילו תלויים במיקום המאורעות בשלבי הביניים של החיפוש. לצורכי החישוב נניח כי במצב התחלתי המאורעות מפוזרים בצורה אחידה על פני הבלוק וכי אחרי כל עידון של מאורע ספציפית היא נשארת במקומה (הנחה הינה סבירה, כי אנחנו מעריכים ממוצע הפעולות.)

יהי c_i מרכזי המאורעות ההתחלתיים השווים ל- $\left\lfloor \frac{N}{M+1} \right\rfloor i,\ 1 \le i \le M$. מספר ההשוואות הממוצע בעידון כל מאורע שווה לאורך של 2 סגמנטים בהם מתבצע החיפוש, כלומר עבור M-1 המאורעות הראשונות נקבל $2\left\lfloor \frac{N}{M+1} \right\rfloor - 1$ השוואות, ועבור המאורע האחרון נקבל $N-(M-1)\left\lfloor \frac{N}{M+1} \right\rfloor - 1$ השוואות. כלומר סה"כ מספר ההשוואות הממוצע הנו:

$$\text{(ii)}\ C_2 = 2(M-1)\left\lfloor \frac{N}{M+1} \right\rfloor + \left(N-(M-1)\left\lfloor \frac{N}{M+1} \right\rfloor\right) - M = \left\lfloor \frac{N}{M+1} \right\rfloor (M-1) + N - M$$

לדוגמא, עבור $N=11,\ M=3$ מקבלים 12 השוואות.

II. מספר החישובים של השגיאה הרגעית

i. ORTD:

<u>חיפוש הסגמנטציה הראשונית</u>: יש לחשב עבור כל מסגרת בבלוק באורך N בכמה סגמנטים שונים היא משתתפת. בהזנחת הקצוות הדבר ניתן לחישוב ע"י:

$$\text{(iii)}\ C_3=\sum_{i=1}^{N} i(N-i)=\sum_{i=1}^{N}\left(iN-i^2\right)=\frac{1}{2}N^2(N+1)-\frac{1}{6}N(N+1)(2N+1)=\frac{N(N+1)^2}{6}$$

<u>חיפוש הסגמנטציה בהינתן וקטורי המטרה</u>: אם ידועים וקטורי המטרה, כל מסגרת עלולה להימצא ב-M בלוקים שונים (בהזנחת הקצוות), לכן ישנם כ- $C_4=MN$ חישובים של שגיאה רגעית. עבור $N=11,\ M=3$, מקבלים $C_4=33,\ \ C_5=264$

ii. SORTeD:

<u>חיפוש הסגמנטציה הראשונית</u>: בכל הסט של מרכז המאורע יש לעשות $2\left\lfloor\frac{N}{M+1}\right\rfloor-1$ חישובים של השגיאה הרגעית בממוצע, כלומר מקבלים מס' פעולות משוערך:

$$C_5=\left(2\left\lfloor\frac{N}{M+1}\right\rfloor-1\right)^2 M \qquad \text{(iv)}$$

<u>חיפוש הסגמנטציה בהינתן וקטורי המטרה</u>: אם ידועים וקטורי המטרה, מספיק לחשב פעם אחת את כל השגיאות הרגעיות, ואז בכל שינוי מיקום מאורע רק 2 מסגרות הקצה ישנו את ערך השגיאה הרגעית שלהן. ניתן לכן להגיע לשגיאה מצטברת חדשה ע"י החסרת שגיאות קודמות והוספת שגיאות חדשות. כלומר, מבצעים כ- $C_6=N+2\left(2\left\lfloor\frac{N}{M+1}\right\rfloor-1\right)$ חישובים של שגיאה רגעית.

עבור $N=11,\ M=3$, מקבלים $C_6=17,\ \ C_5=27$

Bibliography ביבליוגרפיה

1. L. R. Rabiner, R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
2. A. Gersho, R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publisher, Boston, 1992
3. A. M. Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, Wiley Publishers, 1999.
4. W.B. Kleijn, K.K. Paliwal, *Speech Coding And Synthesis*, Elsevier, 1995.
5. T.E. Tremain, "The Government Standard Linear Predictive Coding Algorithm. LPC-10", *Speech Technology*, pp. 40-49, April 1982.
6. A.V. McCree, T.P. Barnwell III, "A Mixed Excitation LPC Vocoder for Low Bit Rate Speech Coding", *IEEE Trans. on Speech and Audio Proc.*, Vol. 3, No. 4, July 1995, pp. 242-250.
7. "The 1200 and 2400 Bits/s NATO Interoperable Narrow Band Voice Coder", NSA, STANAG No. 4591.
8. *Information Technology – Very Low Bitrate Audio-Visual Coding*, International standard, Part 3, Subpart 2, (ISO/IEC FCD 14496-3 Subpart 2, ISO/JTC 1/Sc 29/WG11).
9. F.K. Soong, B.H. Juang, "Line Spectrum Pair (LSP) and Speech Data Compression", *Proc. ICASSP-84*, 1984, pp 1.10.1-1.10.4.
10. H.B. Choi, W.T.K. Wong, B.M.K. Cheetham, C.C. Goodyear, "Interpolation of spectral information for low bit rate speech coding", *Proc. of the European Conf. on Speech Comm. and Tech.*, 1995, pp-1033-1036.
11. K.K. Paliwal "Interpolation properties of linear prediction parametric representation", *Proc. of the European Conf. on Speech Comm. and Tech.*, 1995, pp-1029-1032.
12. Y. Linde, A. Buzo, R.M. Gray, "An algorithm for vector quantization design", *IEEE Trans. Commun.*, vol. COMM-28, pp. 84-95, 1980.
13. K.K. Paliwal and B.S. Atal "Efficient vector quantization of LPC parameters at 24 bits/frame", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* Detroit, MI, 1995, pp 740-743.
14. P. Hedelin, "Single stage spectral quantization at 20 bits", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proc.*, 1994, pp. 1525-1528.
15. J. Makhoul, S Roucos, H. Hish, "Vector Quantization in Speech Coding", *Proc. IEEE*, vol. 73, pp. 1551-1588, Nov. 1985.
16. D.Y. Wong, B.H. Juang, A.H. Gray "An 800 bit/s vector quantization LPC vocoder", *IEEE Trans. on ASSP,* Vol. ASSP-30, No. 3, 1982, pp. 770-780.

17. J. Foster, R.M. Gray, M.O. Dunham, "Finite-state vector quantization for waveform coding", *IEEE Trans. on Inf. Theory*, Vol. 31, pp. 348-359, 1985.

18. Y. Hussain, N. Farvardin, "Finite-state vector quantization over noisy channels and its application to LSP parameters", *Proc. IEEE ICASSP-92*, 1992, vol. 2, pp 133-136.

19. T. Eriksson, J. Linden, J. Skoglund, "Interframe LSF Quantization for Noisy Channels",*IEEE Trans. on SAP,* Vol. 7, No. 5, Sept. 1999, pp. 495-509.

20. Y. Shoham, "Vector predictive quantization of the spectral parameters for low bit rate speech coding", *Proc. ICASSP-87*, Vol.4, pp. 2181-2184.

21. M.Yong, G.Davidson, A.Gersho, "Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* 1988, pp 402-405.

22. C. Tsao, R.M. Gray, "Matrix Quantizer Design for LPC Speech Using the Generalized Lloyd Algorithm", *IEEE Trans. on ASSP,* Vol. ASSP-32, No. 3, June 1985, pp. 537-545.

23. C. S. Xydeas, C. Papanastasiou, "Split Matrix Quantization of LPC Parameters", *IEEE Trans. on SAP,* Vol. 7, No. 2, March 1999, pp. 113-125.

24. S. Bruhn, "Matrix Product Quantization for Very-low-rate Speech Coding", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* Detroit, MI, 1995, pp 724-727.

25. H.P. Knagenhjelm, W.B. Kleijn, "Spectral dynamics is more important than spectral distortion", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* Detroit, MI, 1995, pp 732-735.

26. R.M. Schwartz, S.E. Roucos, "A Comparision of Methods for 300/400 b/s Vocoders", *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.",* 1983, pp 69-72.

27. D. Kemp, J. Collura, T. Tremain, "Multi-frame coding of LPC parameters at 600-800 bps.", *IEEE ICASSP-91*, 1991, pp. 609-612.

28. J.M. Lopez-Soler, V.Sanchez, A. de la Torre, A.J. Rubio-Ayuso, "Linear inter-frame dependencies for very low bit-rate speech coding", *Speech Communication* 34, 2001, pp 333-349.

29. J.M. Lopez-Soler, N. Farvardin, "A combined quantization-interpolation scheme for very low bit rate coding of speech LSP parameters", 1993

30. R. Mayrench, D. Malah, "Low-bit-rate Speech Coding Using Quantization of Variable Length Segments", *Eurospeech-1999.*

31. R. Mayrench, "Low-bit-rate Speech Coding Using Joint Segmentation and Vector Quantization", M. Sc. Thesis, Technion IIT, Haifa.

32. S. Roucos, R. Schwartz, J. Makhoul, "Segment quantization for very low bit rate speech coding", *Proc. IEEE ICASSP-83,* 1983, pp. 1565-1568.

33. S. Roucos, A. Wilgus, W. Russel, “A Segment Vocoder Algorithm for Real-Time Implementation”, IEEE 1987.

34. Y. Shiraki, M. Honda, "LPC Speech coding based on variable length segment quantization", *IEEE Trans. on ASSP,* Vol. 36, Sept. 1988, pp. 1437-1444.

35. M. Honda, Y. Shiraki, "Very low-bit-rate speech coding", in *Advances in Speech Signal Processing*, S.Furui and M.M. Sondhi, Eds. New York: Marcel Dekker, 1992, pp. 209-230

36. P.Prandoni, M. Vetterli, "R/D Optimal Prediction", *IEEE Trans. on Speech and Audio Proc.*, Vol. 8, No. 6, 2000, pp. 646-655.

37. B. Atal, "Efficient coding of LPC parameters by temporal decomposition", *IEEE ICASSP-83*, Boston, pp 81-84.

38. M. Niranjan, F. Fallside, "Temporal decomposition: A Framework for Enhanced Speech Recognition", 1989.

39. A.C.R. Nandasena, P.C. Nguyen, M. Akagi, "Spectral Stability Based Event Localizing Temporal Decomposition", *Computer Speech and Language*, 2001, Vol. 15, no. 4, pp 381-401.

40. S.J. Kim, S. Lee, W.J. Han, Y.H. Oh, "Efficient Quantization of LSF Parameters Based on Temporal Decomposition", ICSLP-1999

41. S.J. Kim, Y.H. Oh "Efficient Quantization Method for LSF Parameters Based on Restricted Temporal Decomposition", *Electronics Letters*, 1999, 10th June, Vol. 35, No. 12, pp. 962-964.

42. P.C. Nguyen, M. Akagi, "Limited error based event localizing temporal decomposition", *Proc. EUSIPCO 2002,* September 2002.

43. P.C. Nguyen, M. Akagi, "Improvement of the Restricted Temporal Decomposition Method for Line Spectral Frequency Parameters, Proc. ICASSP-2002.

44. C.N. Athaudage, A.B. Bradley, M. Lech, "Optimization of Temporal Decomposition Model of Speech", *Proc. on 5th Int. Symp. on Signal Proc. and its Applications, ISSPA '99*, Brisbane, Australia, 1999, pp. 471-474.

45. C.N. Athaudage, A.B. Bradley, M. Lech, "Model-Based Speech Signal Coding Using Optimized Temporal Decomposition for Storage and Broadcasting Applications", *EURASIP Journal On Applied Signal Processing,* 2003, Oct., pp 1016-1026.

46. Y.M. Cheng "Short-Term Temporal Decomposition and its Properties for Speech Compression", *IEEE Trans. on Signal Proc.*, Vol. 39, No. 6, June 1991, 1282-1290.

47. A.M.L.V. Dijk-Kappers, S.M. Marcus, "Temporal decomposition of speech", *Speech Communication*, Vol. 8, No. 2, pp. 125-135.

48. S. Ghaemmaghami, M. Deriche, “A new approach to very low-rate speech coding using temporal decomposition”, IEEE-1996

49. S. Ghaemmaghami, M. Deriche, “Adaptive-width approximation of events in temporal decomposition based speech coding”, *Electronics letters*, Vol. 32, No. 24, Nov. 1996, pp. 2189-2191.

50. S. Ghaemmaghami, M. Deriche, B. Boashash, "Comparative Study of Different Parameters for Temporal Decomposition Based Speech Coding", IEEE-1997

51. Y. Cheng, D. O'Shaughnessy, "On 450-600 b/s natural sounding speech coding", *IEEE Trans. on Speech and Audio Proc.*, Vol. 1, 1993, pp. 207-220.

52. S. Ghaemmaghami, M. Deriche, “A new approach to efficient interpolative determination of pitch contour using temporal decomposition", *IEEE TENCON, DSP Applications*, 1996, pp 125-130.

53. P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound", *Institute of Phonetic Sciences, University of Amsterdam, Proceedings*, 17 (1993), pp. 97-110.

54. F. Itakura, “ Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals”, *J. Acoustic Society of America*, 1975.

55. W.R. Gardner and B.D. Rao, "Theoretical Analysis of the High Rate Vector Quantization of LPC Parameters", *IEEE Trans. Speech, Audio Processing*, Vol. 3, No. 5, pp. 367-381, 1995.

56. J. Samuelsson, J. Skoglund, and J. Lindén, "Controlling Spectral Dynamics in LPC quantization for perceptual enhancement", *Proc. 31st Asilomar Conference on Signals, Systems, and Computers*, 1997, pp. 1066-1070, Pacific Grove, CA.

57. Mark W. Chamberlain, "A 600 bps MELP vocoder for use on HF channels", *MILCOM 2001 – IEEE Military Communications Conference*, no. 1, October 2001 pp. 447-453

58. NAG Fortran Library Manual, Mark 20, The Numerical Algorithm Group Limited, 2002

59. W. H. Press, Numerical Recipes in C, CH. 2, Cambridge University Press 1992.

60. Rix, A. W., Beerends, J. G., Hollier, M. P. and Hekstra, A. P. "Perceptual Evaluation of Speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs". *ITU-T recommendation P.862*, February 2001

61. A. McCree, J. C. De Martin, “A 1.7 KB/S MELP Coder with Improved Analysis and Quantization”, *IEEE Int. Conference on Acoustics, Speech, and signal Processing, 1998.*

62. W.-J. Han, E.-K. Kim Y.-H. Oh, "Natural quality two band LPC coding of speech at 880 bit/s with frame interpolation", *Electronic Letters* , 2002, Vol. 38, No. 6, pp 292-294.

Very Low Bit-Rate Speech Coding based on Temporal Decomposition

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

SLAVA SHECHTMAN

SUBMITTED TO THE SENATE OF THE
TECHNION – ISRAEL INSTITUTE OF TECHNOLOGY

AV 5764 HAIFA JULY 2004

To Beloved Nadusha and Talik

THE RESEARCH THESIS WAS DONE UNDER THE SUPERVISION OF PROFESSOR DAVID MALAH IN THE FACULTY OF ELECTRICAL ENGINEERING

I wish to thank Professor David Malah for his devoted and thorough guidance, substantial involvement and the fruitful discussions.

I would like to thank the Signal and Image Processing laboratory staff: Nimrod Peleg, Yair Moshe, Ziva Avni, Avi Rozen and Tamara Gvirtz for their help and technical support.

Thanks for Guy Narkiss, Tzachi Zehavi and Asaf Hargil for the implementation and the provision of output examples of the system, which served as a main comparison point for the algorithm that was developed in the current work.

I wish to thank my wife Nadya for her love, encouragement and eternal patience, my wife's and my parents for their belief in me and for their support along the way.

The generous financial help of the Technion is gratefully acknowledged.

Table of Contents

Abstract **1**

List of symbols **3**

Chapter 1. Introduction **6**

Chapter 2. Background for speech coding systems **10**

2.1 The structure of speech signal 10

2.2 The structure of parametric voice coder 11

2.3 LPC model for spectral envelope representation 11

2.3.1 The model 11

2.3.2 Spectral envelope estimation by linear prediction 13

2.3.3 LPC filter representation by LSF 14

2.4 Distortion functions for the LPC 15

2.4.1 The Log Spectral Distortion (LSD) 16

2.4.2 WMSE based error criteria 16

2.5 Mixed Excitation Linear Prediction (MELP) vocoder 18

Chapter 3. Methods for compression of spectral parameter vectors **21**

3.1 Introduction 21

3.2 Spectral parameter coding 21

3.3 Low delay temporal redundancy removal 22

3.4 Selected frame transmission 23

3.5 Matrix Quantization (MQ) 23

3.6 Segment Quantization (SegQ) 25

3.7 Segment representation with varying basic analysis frame length 27

3.8 Summary 29

Chapter 4. Temporal Decomposition (TD) model 30

4.1 Introduction 30

4.2 TD Generic model 30

4.2.1 General description 30

4.2.2 Target matrix calculation 33

4.3 SVD based TD method 33

4.4 Modern TD techniques, assuming initial segmentation 36

4.4.1 Soft restriction of event function support by Lagrange multipliers 37

4.4.2 Hard restriction of event function support (RTD) 39

4.5 Optimal Restricted Temporal Decomposition (ORTD) 43

4.5.1 General description 43

4.5.2 Block overlapping and edge effects 45

4.6 Summary 46

Chapter 5. Spectral envelope coding by DW-SORTeD 47

5.1 Introduction 47

5.2 Error criteria for spectral envelope modeling 48

5.2.1 General 48

5.2.2 A filtered G-WSE criterion 48

5.3 Dynamically Weighted Optimal RTD (DW-ORTD) 50

5.3.1 The event function determination stage modification 50

5.3.2 The target refinement stage modification 50

5.4 Sub-Optimal Resctricted Temporal Decomposition (SORTeD) 53

5.4.1 Introduction 53

5.4.2 General Description 54

5.4.3 Analysis block update 55

5.4.4 Initial segmentation determination 56

5.4.5 Suboptimal search algorithm for the initial segmentation improvement 59

5.4.6 Complexity comparision of SORTeD vs. ORTD 62

5.4.7 Instantaneous event function calculation 64

5.5 Quantization of TD parameters. 70

5.6 Summary 72

Chapter 6. Performance estimation of DW-SORTeD 73

6.1 Introduction 73
6.2 Performance estimation of TD model 74
6.2.1 General 74
6.2.2 ORTD vs. DW-ORTD with different dynamic weighting 74
6.2.3 Parameter examination of DW-SORTeD 76
6.2.4 DW-SORTeD vs. DW-ORTD 77
6.2.5 Constrained DW-SORTeD 78
6.2.6 DW-SORTeD performance convergence 79
6.3 Performance evaluation of TD model with quantization 80
6.3.1 Target vector codebook comparision 81
6.3.2 Event fumction coding performance 81
6.3.3 DW-SORTeD performnace at different rated and delays 82
6.4 Summary 85

Chapter 7. DW-SORTeD for excitation parameter compression 86

7.1 Introduction 86
7.2 Constrained RTD suitability for excitation parameters compression 87
7.2.1 Introduction 87
7.2.2 Constrained RTD under affine transform 88
7.3 TD model for joint excitation parameter compression 89
7.3.1 General description 89
7.3.2 Preprocessing of the excitaiton parameters 90
7.3.3 Normalization 91
7.3.4 Dynamic weighting for excitation parameters 93
7.3.5 DW-SORTeD adjustment for excitation parameter coding 95
7.4 Performance estimation of DW-SORTeD for excitation parameter coding. 95
7.4.1 General 95
7.4.2 Run example 95

VLBR coder based on DW-SORTeD and MELP 99

8.1 General 99

8.2 The MELP based 600-650 bps coder 99

8.2.1 General 99

8.2.2 Voicing pattern quantizaion 100

8.2.3 Bit allocation examples for 600-650 bps coders 102

8.2.4 Bit allocation for 800-880 bos coders 104

8.3 Performance estimation 105

8.3.1 Coding schemes to compare 105

8.3.2 The comparision results 106

8.3.3 Discussion and conclusions 107

Summary and suggestions for further research 109

9.1 Summary 109

9.2 Suggestions for further research 111

Appendix I 112

Appendix II 114

Bibliography 116

List of tables and figures

List of Tables

Table 2-a. Transparent quality of spectral envelope as defined by the LSD measure....... 16
Table 2-b. MELP coder bit allocation for a single frame 20
Table 5-a. Initial segmentations, examined for the SORTeD algorithm.............................58
Table 5-b. Sub-optimal algorithm for block segmentation search (inside SORTeD). The initial application 60
Table 5-c. Sub-optimal algorithm for block segmentation search (inside SORTeD). Non-initial run 61
Table 5-d. Suboptimal (SORTeD) vs. optimal (ORTD) segmentation complexity 63
Table 6-a. Spectral distortion (average LSD and outliers percentage), obtained for different weighting of DW-ORTD error criterion..76
Table 6-b. Experimental results for examining different initial segmentation setups of DW-SORTeD. 76
Table 6-c. Spectral distortion (average LSD and outliers percentage), obtained for DW-ORTD and DW-SORTeD. 77
Table 6-d. Performance of the constrained DW-SORTeD algorithm compared to appropriate unconstrained algorithm. 79
Table 6-e. Performance of different configurations of constrained DW-SORTeD algorithm with modified Gardner weights..79
Table 6-f. Performance of different codebooks for LSF quantization by Split-VQ. 81
Table 6-g. . Performance of DW-SORTeD with unquantized target vectors 82
Table 6-h. The evaluation of DW-SORTeD for spectral envelope coding. 84
Table 7-a. Performance of excitation quantization by Temporal Decomposition. 97
Table 8-a. Quantization table of voicing parameters. 102
Table 8-b. . Bit allocation for MELP-based speech coder with 11-frames block 103
Table 8-c. . Bit allocation for MELP based speech coder with 7-frames block 104
Table 8-d. Performance estimation of VLBR coders. 106

List of Figures

Figure 2-1 A generic parametric voice coder system 11
Figure 2-2. The basic LPC Model 12
Figure 3-1. Forward-Backward PVQ, as used in MELP-1200 standard for LSF quantization 22
Figure 3-2. Finding J segments by dynamic programming. 26
Figure 3-3. Trellis diagram of the dynamic programming algorithm. 29
Figure 4-1. Temporal Decomposition notation. 31
Figure 4-2. General TD stages. 32
Figure 4-3. General Temporal Decomposition model. 32
Figure 4-4. Weighting function for the *k*-th event 37
Figure 4-5. Typical shape of an initial event function. 38
Figure 4-6. Restricted Temporal Decomposition (RTD) model of speech. 39
Figure 4-7. RTD with optimal (non-constrained) event functions. 40
Figure 4-8. Optimal instant event function scatter for RTD model. 41
Figure 4-9. RTD with one's complementary event functions. 42
Figure 4-10. A trellis example for best segmentation search in ORTD. 44
Figure 4-11. Buffer overlap in ORTD. 46
Figure 5-1. LSD matching ability of weighted squared errors. 49
Figure 5-2. Sub Optimal RTD (SORTeD) algorithm. General description 55
Figure 5-3. Analysis block overlap for constant rate scheme. 56
Figure 5-4. RTD with one's complementary non-negative and centered event functions.. 66
Figure 5-5. RTD with one's complementary, non-negative, centered and monotonic event functions branches. 68
Figure 5-6. Improved monotony constraint. 70
Figure 6-1. Spectral distortions (average LSD), obtained for different weighting of DW-ORTD's error criterion 75

Figure 6-2. Spectral distortions (average LSD), obtained for DW-ORTD and DW-SORTeD....78
Figure 6-3. Improvement of DW-SORTeD performance vs. number of its iterations....80
Figure 6-4. The evaluation of DW-SORTeD for spectral envelope coding by average LSD scores....84
Figure 6-5. The evaluation of DW-SORTeD for spectral envelope coding, combined with the MELP standard excitation, using PESQ scores....85
Figure 7-1. Pitch-Energy joint trajectory in time....87
Figure 7-2. Excitation TD....90
Figure 7-3. Minimum and maximum energy tracking for excitation parameter vector normalization....92
Figure 7-4. Average pitch frequency tracking for excitation parameter vector normalization....92
Figure 7-5. Dynamic energy weighting for joint pitch-energy quantization with DW-SORTeD algorithm....94
Figure 7-6. Dynamic pitch weighting for joint pitch-energy quantization with DW-SORTeD algorithm....94
Figure 7-7. Joint DW-SORTeD of pitch and energy.....96
Figure 8-1. Block diagram of 600 bps vocoder....100
Figure 8-3. Performance estimation of VLBR coders....107
Figure 8-4. Perceptual quality degradation for different VLBR algorithms....108

Abstract

This work deals with very low bit-rate (VLBR) speech coding using Temporal Decomposition (TD) techniques. TD is a method of modeling a set of consecutive speech parameter vectors as a sequence of stable event parameter vectors (or targets) and an associated set of overlapping interpolation functions (event functions), centered at the corresponding event instants. The TD technique serves for removing temporal redundancy from the sequence of speech spectral envelope vectors.

An algorithm for efficient representation and coding of the spectral envelope parameters (the Line Spectral Frequencies), based on TD concepts, is proposed. First, this modeling technique was incorporated with quantization to obtain an about 300 bps spectral envelope coder, gaining perceived output speech quality comparable to 1100 bps envelope coding of a standard MELP-2400 codec. Then, the same technique assisted to jointly quantize excitation parameters (pitch and gain) with a bit rate under 300 bps, thus resulting in a full speech coder working at about 600 bps. The coder was based on MELP-2400 standard parameter extraction.

The proposed algorithm is based on the Optimized Restricted-Temporal-Decomposition (ORTD) technique for speech envelope representation, under a MMSE criterion [45]. This algorithm imposes that only two adjacent event functions may overlap, thus allowing a closed analytic solution for instantaneous event functions, given targets and event instants. The ORTD is applied on a block of parameter vectors, and it performs a full search for all possible event instants placements, assuming a given number of events inside the block. The event functions determination stage is followed by a target refinement stage. Both stages alternate until the solution converges. The ORTD has been shown to be promising for very low bit rate speech coding, but only for storage and broadcast applications, due to its high computational load.

In order to improve perceptual speech quality of the ORTD, a dynamically weighted ORTD (DW-ORTD) technique is introduced in this work. It extends the ORTD by allowing temporally changing weights, so as to improve the perceived speech quality. Using a modified version of Gardner's weighted MSE with DW-ORTD is found to

provide a reduction in the Log Spectral Distance (LSD) measure by 0.3 dB, as compared to ORTD (from 1.75 dB to 1.4 dB for 11.75 events/second).

The original ORTD algorithm delay and complexity requirements make it inappropriate for real-time speech coding. In this work we also introduce a modification of this technique, denoted Sub-Optimal RTD or SORTeD, which is sub-optimal but is suitable for on-line speech coding purposes. The sub-optimal solution performs a partial search over possible event instants, dependant on their initial placement. Appropriate setting of initial conditions, integrated with block edge effects removal by adjacent block overlapping resulted in a negligible degradation of the sub-optimal algorithm as compared to the optimal one, proposed in ORTD. The degradation of performance, in terms of LSD, was only about 0.06 dB.

The direct solution for determining instantaneous event functions may result in irregular shapes, which could be difficult to quantize using vector quantization. It has been proposed to regularize those functions, by imposing constraints on the optimal ORTD solution, such as the one's complement property of each event function pair at each time instant, in addition to non-negativity and monotonicity of each event function throughout a segment (i.e. the interval between adjacent event instants).

The algorithm, combining both techniques (and denoted DW-SORTeD), was incorporated with target vectors and event functions quantization to obtain a very low bit rate spectral envelope coding scheme. The parameters to be quantized are the target vectors and the event functions. For the constrained event functions, only the decreasing branch shape and its length have to be quantized. Shape quantization, that fits the very-low-bit-rate paradigm, is performed by a small codebook VQ of decreasing event functions branches. For each possible length, a different codebook is stored, while the allowable distance between adjacent event instances is limited to 8 (3 bit length code). The targets are quantized, as common, by the Split-VQ technique of the lower 4 and upper 6 LSF parameters. Small event-function shape codebooks (2-4 bits) are trained with event functions created by the constrained Dynamically Weighted SORTeD (DW-SORTeD) algorithm.

The DW-SORTeD scheme, combined with quantization, allows representation of the speech spectral envelope by as low as 300 bps (constant rate), still preserving on acceptable speech quality and imposing only moderate computational requirements. The algorithmic delay of the encoder (i.e. the block length to minimize an error criterion on) is 11 frames or 7 frames, with a frame duration of 22.5 ms (thus resulting in about 250 ms or 160 ms of algorithmic delay). The quality of the 7-frame-delay coding scheme is inferior compared to the 11-frame-delay scheme, and about 80 bps more are needed for the 7-frames delay coder to reach the 11-frame-delay system performance.

To assess the spectral envelope coding subjective quality, the DW-SORTeD has been incorporated into a MELP-2400 standard codec (thus presenting a 1.6 Kbps codec). Its objective quality was estimated with the PESQ (ITU P.862) standard for an objective evaluation of the perceptual quality, showing a degradation of 0.18 – 0.25 of the estimated MOS score (i.e. PESQ score), as compared to MELP-2400 spectral quantization (2.75-2.82 instead of 3.0).

In order to reduce the bit rate by the MELP standard for the excitation parameters, their number was reduced, and only pitch, gain and voicing were left. Pitch and gain were jointly coded by DW-SORTeD (dynamic weights for error criterion and prior normalization made it possible), achieving a bit-rate as low as 260-340 bps for the excitation parameters. The four-band-voicing information (as defined in the MELP standard) was roughly quantized to 2 bits per frame and then decimated in time, adding about 25 bps to the overall bit-rate.

The full speech coder, featuring 600-650 bps and 11 frames of algorithmic delay, gained 2.6-2.65 of PESQ score, respectively. A similar system with 7 frames of algorithmic delay gained 2.5-2.6 for 610-650 bps' accordingly. Those results were compared to a previously reported 600 bps system [57], based also on MELP-2400 standard, and performing joint quantization of 4 consecutive frames of speech, and revealed an improvement of 0.2-0.3, over that coder, in terms of PESQ estimation of MOS score.